

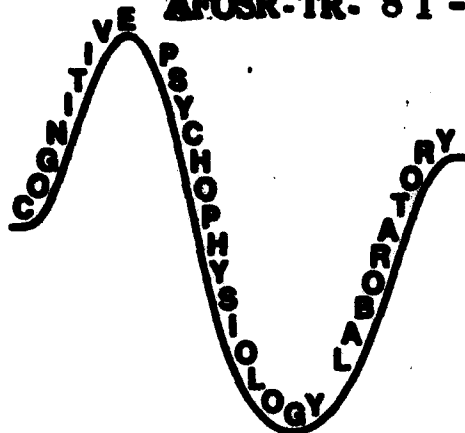
AD-A097 542 ILLINOIS UNIV CHAMPAIGN COGNITIVE PSYCHOPHYSIOLOGY LAB F/G 5/9  
APPLICATIONS OF EVENT-RELATED BRAIN POTENTIALS IN HUMAN ENGINEE--ETC (U)  
NOV 80 E DONCHIN, C D WICKENS, R L HORST F49620-79-C-0233  
UNCLASSIFIED CPL80-1/AFO-APP AFOSR-TR-81-0306 NL

1 of 1  
25  
2007-04-24



END  
DATE  
FILMED  
5 81  
DTIC

AD A 097542



AFOSR-TR-81-0806

**LEVEL III**

(12)

Department of Psychology  
University of Illinois  
Champaign, Illinois 61820

Technical Report No. CPL88-1/APO  
SR8-1

November 1980

**APPLICATIONS OF EVENT RELATED  
BRAIN POTENTIALS  
IN HUMAN ENGINEERING  
ANNUAL PROGRESS REPORT: *1987-007*  
(Appendices) *TR-81-0345***

**EMANUEL DONCHIN AND CHRISTOPHER WICKENS  
COGNITIVE PSYCHOPHYSIOLOGY LABORATORY**

**DTIC  
ELECTE  
APR 2 1981**

**S D**

Prepared for:  
The Air Force Office of Scientific Research  
1215 Research Building  
AFOSR-TR-81-0806  
November 1980

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

| REPORT DOCUMENTATION PAGE  |                       | READ INSTRUCTIONS<br>BEFORE COMPLETING FORM                 |
|--|-----------------------|---|
| 1. REPORT NUMBER   | 2. GOVT ACCESSION NO. | 3. RECIPIENT'S CATALOG NUMBER                               |
| AFOSR-TR-81-0306   |                       | AD-A097542  |
| 4. TITLE (and Subtitle)  |                       | 5. TYPE OF REPORT & PERIOD COVERED                          |
| Applications of Event-Related Brain Potentials in Human Engineering. Annual Progress Report, Appendices.   |                       | INTERIM   |
| 7. AUTHOR(s)   |                       | 6. PERFORMING ORG. REPORT NUMBER                            |
| Emanuel Donchin and Christopher D. Wickens<br>Richard L. Horst Emanuel   |                       |   |
| 8. CONTRACT OR GRANT NUMBER(s)   |                       |   |
| 149629-79-C-0233   |                       |   |
| 9. PERFORMING ORGANIZATION NAME AND ADDRESS  |                       | 10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS |
| University of Illinois, Department of Psychology<br>Champaign, IL 61820  |                       | 61102F 17 1-1<br>2313/A4                                    |
| 11. CONTROLLING OFFICE NAME AND ADDRESS  |                       | 12. REPORT DATE   |
| Air Force Office of Scientific Research (NL)<br>Bolling AFB DC 20332   |                       | 11 NOV 80 (12 92)   |
| 14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office)  |                       | 13. NUMBER OF PAGES   |
| (14) CP 80-1/HFB-A1-E<br>SF 8-1-HAP  |                       | 100   |
| 15. SECURITY CLASS. (of this report)   |                       | 15a. DECLASSIFICATION/DOWNGRADING SCHEDULE                  |
| Unclassified   |                       |   |
| 16. DISTRIBUTION STATEMENT (of this Report)  |                       | Accession For   |
| Approved for public release; distribution unlimited.   |                       | NTIS GRA&I <input checked="" type="checkbox"/>              |
|  |                       | DTIC TAB <input type="checkbox"/>                           |
|  |                       | Unannounced <input type="checkbox"/>                        |
|  |                       | Justification <input type="checkbox"/>                      |
| 17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)   |                       | By <input type="checkbox"/>                                 |
|  |                       | Distribution <input type="checkbox"/>                       |
|  |                       | For <input type="checkbox"/>                                |
|  |                       | Disc <input type="checkbox"/>                               |
| 18. SUPPLEMENTARY NOTES  |                       | A   |
| Proceedings of the Human Factors Society -<br>24th Annual Meeting, 1980  |                       |   |
| 19. KEY WORDS (Continue on reverse side if necessary and identify by block number)   |                       |   |
| Event-related brain potentials, P300, attention, workload, expectancy, subjective probability  |                       |   |
| 20. ABSTRACT (Continue on reverse side if necessary and identify by block number)  |                       |   |
| This report describes research partially or entirely conducted during the first year of the contract period. It describes experiments related to four basic categories of human performance research: (1) attention (including both attention allocation and workload), (2) subjective probability and expectancy, (3) processing latency, and (4) control movement. A series of appendices describes those portions of the research that have been completed. |                       |   |

DD FORM 1 JAN 73 1473

EDITION OF 1 NOV 85 IS OBSOLETE  
S/N 0103-014-6601

UNCLASSIFIED 42249  
SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

Appendix A

THE EVENT-RELATED BRAIN POTENTIAL AS AN INDEX OF ATTENTION ALLOCATION  
IN COMPLEX DISPLAYS

Christopher D. Wickens, Earle F. Heffley,  
Arthur F. Kramer, and Emanuel Donchin

Cognitive Psychophysiology Laboratory  
Department of Psychology  
University of Illinois  
Champaign, Illinois 61820

ABSTRACT

The advantages of employing the event-related brain potential (ERP) in the assessment of allocation of attention in dynamic environments are discussed. Three experiments are presented in which the P300 component of the ERP is demonstrated to be a useful index of subjects' locus of attention. The first two experiments were concerned with the allocation of attention during discrete and continuous visual monitoring tasks. The results indicated that a P300 was elicited only by stimuli to which the subject had to attend in order to perform successfully the task. The third experiment was conducted to assess the sensitivity of P300 to the manner in which attention is allocated to different aspects of a display during the performance of a 3-dimensional target acquisition task. The amplitude of the P300 was found to reflect differences between two levels of workload, as well as the task relevance of the stimuli. The results of the experiments are discussed in terms of their utility in the evaluation of the design of man-machine systems as well as in the study of the allocation of attention in operational environments.

The manner in which the event-related brain potential (ERP) may be used in the assessment of workload imposed upon the operators of man-machine systems has been described previously (Isreal, Chesney, Wickens & Donchin, 1980; Isreal, Wickens, Chesney & Donchin, 1980; Isreal, Wickens & Donchin, 1979; Wickens, Isreal & Donchin, 1977). Specifically, the amplitude of the late positive (P300) component of the ERP has been shown to vary as a function of the perceptual demands imposed upon the operator. The experiments described in this report illustrate how the amplitude of P300 can be used to ascertain the operator's allocation of attention to different aspects of a complex display.

The ERP based determination of the focus of attention can be of value in two contexts. In an off-line context, ERPs can be used during the system's design. An evaluation of the degree to which attention is allocated to different information sources can allow the system designers to highlight important, but neglected, channels and deemphasize irrelevant channels that attract unnecessary attention. In an on-line context, monitoring the allocation of attention may enhance the effectiveness of adaptive systems. For example, when the ERP indicates that the operator has failed to detect a warning signal, an adaptive system could act to increase the

salience of the signal. Rouse (1977) has argued for the advantage of cooperative man-machine system interaction, in which a computer's knowledge of the tasks the operator is performing at each moment enables the computer to assume responsibility for neglected activities.

In multi-operator systems, there appears to be merit in a system that can alert other operators when it determines that an operator has failed to attend to important information. Weiner (1977), in his analysis of controlled flight into terrain, makes the telling point that the crash of Eastern Airlines Flight 401 into the Florida Everglades might have been averted, had the air traffic controller known that no one on the flight deck had noticed the ground proximity warning.

We emphasize that the information the ERP can offer concerning resource allocation is complementary to the data that is derived from traditional sources. The primary advantage of the ERP is that it does not require the operator to make an overt response to the eliciting stimuli. Thus, the ERP may be useful to index the allocation of attention when the operators are monitoring displays. Other non-invasive physiological techniques exist but they have serious limitations. Autonomic measures are commonly dissociated from stimulus

processing under the conditions of workload and stress found in many man-machine interactions. Although ocular fixation measures also do not require overt responses, the direction of gaze need not correspond to the channel being processed. Eye tracking systems are particularly ineffective when multiple channels are present in the fovea or when signals are delivered over auditory channels.

Donchin and Cohen (1967) and Eason and Ritchie (1977) have shown that the P300 elicited by attended stimuli in a visual array is larger than that elicited by unattended stimuli. Because these investigations used relatively simple stimuli and tasks, it is not clear that the results can be generalized to the complex displays encountered in real systems. The experiments reported below extend these results to displays which are more similar to displays monitored by the operators of contemporary man-machine systems.

#### EXPERIMENT 1 MONITORING DISPLAYS OF VARYING COMPLEXITY

The first experiment in this series required subjects to monitor a simulated air traffic control display. Subjects watched a display screen on which several squares and triangles (0.4 X 0.4 cm) appeared along one edge and then traversed the screen in a linear path. The subjects were instructed to pay attention to the squares and to ignore the triangles. A square or a triangle was briefly intensified every few seconds. We call these intensifications 'flashes'. The subject's task was to monitor the squares and count the number of times they flashed. Each monitoring period lasted four minutes. Display complexity was varied by changing the number of relevant and irrelevant elements (squares/triangles) on the screen during different monitoring periods.

Two aspects of the results, described by Heffley, Wickens and Donchin (1978), are important. Relevant flashes elicited a large P300 in the average ERPs obtained from all subjects at all levels of display complexity. Irrelevant flashes elicited small and inconsistent P300 components. In addition, the latency of P300 to relevant flashes was systematically greater during the monitoring periods in which a larger number of irrelevant or relevant elements appeared on the screen. It should be noted that this increase in P300 latency with display complexity occurred even though the subject's counting did not vary. These data confirm the assertion that the amplitude of the P300 can be used to assess the allocation of attention. The latency appears to be a useful index of display complexity.

For the P300 to be useful in assessing the allocation of attention in real-time, the difference between the ERP elicited by relevant and irrelevant flashes must be detectable following single flashes. The data were therefore subjected to a linear step-wise discriminant analysis (SWDA). This procedure, described by Squires and Donchin (1977), determines the set of independent time points in the waveforms that most clearly distinguished the two classes. The discriminant function classified correctly 83% of the single trial ERPs.

#### EXPERIMENT 2 DISCRETE VERSUS CONTINUOUS MONITORING TASKS

In Experiment 1, we examined P300 in a visual monitoring task which allowed the subject to ignore with ease the irrelevant events. Because the squares and triangles were continuously viewable, the subject could visually track the relevant symbols (squares) and, essentially, filter out the irrelevant elements (triangles).

In Experiment 2, we examined the ERPs elicited by the same stimuli when the subjects are not able to visually track (and thus allocate attention to) the relevant elements. This was accomplished by blanking the entire display except for the brief moment when a flash occurred in one of the elements. At that point, the subject saw the entire set of squares and triangles, one of which was brighter than the others. The subject's task was to identify the bright element and count the bright element if it was a square. Each subject experienced several monitoring periods. In half of the periods the display was on continuously as in Experiment 1. In the other half we used the discrete display.

The data indicate that, when the display does not permit selective attention to focus on a class of events, the irrelevant events evoke P300 components with amplitudes just slightly smaller than those elicited by the relevant events. The increase in the amplitude of P300 following irrelevant events can be observed by comparing the waveforms in Figure 1, top, which represent the average of ERPs for twelve subjects.

These data are summarized in the graph at the bottom of Figure 1. The relevant events elicit larger P300 components whether the display is continuous or discrete. Note that quite a large P300 is elicited by the irrelevant events in the Discrete condition. This striking difference between the response to uncounted events in the two conditions may

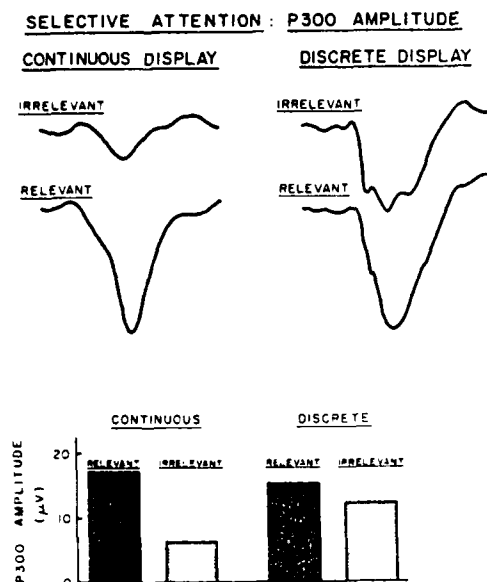


Figure 1. Top: Grand average ERPs (12 subjects) time-locked to intensifications of squares (relevant) and triangles (irrelevant). (Pz electrode; positive downward) Bottom: Mean baseline-to-peak amplitude measures of P300 (averaged across twelve subjects).

be explained if we assume that in the Discrete conditions the subject was forced to actively process both squares and triangles.

### EXPERIMENT 3 DYNAMIC TARGET ACQUISITION

In the preceding experiments the operator was not required to make any overt responses. In Experiment 3, we examine a dynamic environment in which a 3-dimensional target acquisition task is assigned, and its performance requires the operator to manipulate two control devices. Again, we assess the ability of the ERPs to evaluate the allocation of attention between different aspects of a display. The experimental design also allowed an evaluation of the magnitude of workload imposed by the task.

Subjects viewed a CRT display upon which a rotating target traversed the screen. Using a two-axis joystick, the subject initially

brought a cursor into spatial alignment with the target. The subject was then required to match the angular velocities of the target and the cursor while maintaining the spatial match. A single axis joystick was used to perform the angular velocity match. When both spatial and angular velocity were matched within a specified criterion, a capture button was depressed and the trial was terminated. The success or failure of the capture was indicated after each trial.

Workload was varied in two ways. In separate conditions, we used either a 1st order (easy) or a 2nd order (difficult) system. Within each order, the difficulty of the task varied from the first phase, when position only was controlled, to the final orientation phase, when the requirement to control the angular velocity was added.

Every 1.5 sec as they were traversing the screen either the target or the cursor briefly intensified. The subject was instructed, in different conditions, either to count intensifications of the target or of the cursor.

Eight well trained subjects participated in 180 trials over a period of two sessions. None of the subjects had prior experience in tracking or ERP experiments.

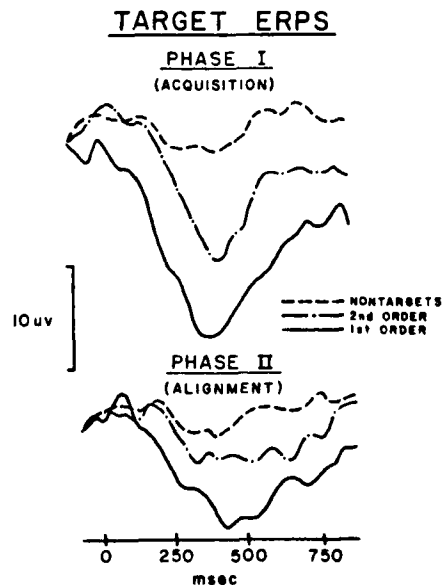


Figure 2. Averaged ERPs recorded from Pz. Each trace represents an average of eight subjects. The top panel displays waveforms for the acquisition phase while the bottom panel presents waveforms for the alignment phase.

The results are presented in Figure 2. Displayed are the ERPs elicited by the counted intensifications during both first (solid) and second (dotted) order tracking along with those elicited by the uncounted intensifications (dashed). The waveforms are presented for the acquisition phase (top panel) and the final alignment phase (bottom panel). Two aspects of the results are of interest:

(1) The P300 component elicited by the counted stimuli is larger in amplitude than the P300 elicited by the uncounted stimulus. This observation replicates, in the more dynamic environment, the effects observed in Experiments 1 and 2.

(2) Both dimensions of workload influenced P300 amplitude. Thus the P300 elicited during second order tracking is consistently of lesser amplitude than the P300 associated with first order, while the overall amplitude of both waveforms is attenuated in the more demanding, final phase of the experimental task. The task difficulty influence on P300 amplitude replicates the prior results of Isreal et al (1980).

#### DISCUSSION

The collective results of these three experiments indicate the robustness of the attention effect. In each experiment stimuli, or events that needed to be attended in order to perform the required task, elicited large P300s. In experiment 1, the triangles, whose lack of relevance could be established by the subject from their spatial location before they flashed (and hence could be ignored), failed to elicit a P300. In experiment 2 the intensifications of both the triangles and squares had to be attended to perform the task, and the P300 was enhanced. In experiment 3, intensifications of whichever task element, target or cursor, the subject was required to process, elicited a P300.

It is instructive to consider why the uncounted stimulus (target or cursor in different conditions) did not elicit a P300 in Experiment 3. This stimulus was clearly necessary if the subject was to perform the tracking task. Both the target and cursor needed to be processed to align the later with the former. Yet, it was only the spatial positions of the target and the cursor that were relevant for tracking; the intensity changes of these stimuli were not relevant to the tracking task. It would appear that the P300 is sensitive to the specific attributes of a stimulus that are utilized in task performance.

Our results have implications for the utilization of ERPs in on-line and off-line systems evaluation. The data suggest that ERPs are useful indices of allocation of attention when a definite task is associated with potentially relevant stimulus events. Thus, the counting task was imposed in order to obtain interpretable P300 components. The counting task may be regarded as one of many possible tasks that require the operator to update an "internal model" of the environment. Other operations sufficient to elicit a P300 may be the acknowledgement of a warning signal, a change-in-status signal, or a verbal command. We are currently investigating the utility of P300 in several, more realistic, monitoring paradigms.

#### ACKNOWLEDGEMENTS

Experiments 1 and 2 were supported by the Advanced Research Projects Agency's Cybernetic Technology Office under ONR Contract #N-000-14-76-C-0002 with Dr. Craig Fields as technical monitor, and by the Air Force Office of Scientific Research under contract F49620-79-C-0233 with Dr. Alfred Fregly as technical monitor. Experiment 3 was supported by a subcontract from NASA - Jet Propulsion Laboratory No. 955610 with Dr. John Hestenes as technical monitor.

#### REFERENCES

- Donchin, E., & Cohen, L. Average evoked potentials and intramodality selective attention. Electroencephalography & clinical neurophysiology, 1967, 22, 537-546.
- Eason, R. G. and Ritchie, G. Effects of stimulus set on early and late components of visually evoked components. Paper presented at the 17th annual meeting of the Psychonomic Society, St. Louis, November 11-13, 1976.
- Heffley, E., Wickens, C., and Donchin, E. Intramodality selective attention and P300 -- Reexamination in a visual monitoring task. Psychophysiology, 1978, 15, 269-270. (Abstract)

Isreal, J. B., Chesney, G. L., Wickens, C. D., and Donchin, E. P300 and tracking difficulty: Evidence for multiple resources in dual-task performance. Psychophysiology, 1980, 17, 259-273.

Isreal, J. B., Wickens, C. D., Chesney, G. L. and Donchin, E. The event-related brain potential as an index of display-monitoring workload. Human factors, 1980, 22, 211-224.

Isreal, J. B., Wickens, C. D., and Donchin, E. P300 amplitude changes during a tracking task as a function of continuous variations in tracking difficulty. Psychophysiology, 1979, 16, 175. (Abstract)

Rouse, W. B. Human-computer interaction in multi-task situations. IEEE transactions in systems, man and cybernetics, 1977, SMC-7, 384-392.

Squires, K. C., & Donchin, E. Beyond averaging: The use of discriminant functions to recognize event related potentials elicited by single auditory stimuli. Electroencephalography & clinical neurophysiology, 1976, 41, 449-459.

Weiner, E. Controlled flight into terrain. Human factors, 1977, 19, 177-182.

Wickens, C. D., Isreal, J. B., and Donchin, E. The event-related cortical potential as an index of task workload. In A. S. Neal and R. F. Palasek (Eds.), Proceedings of the Human Factors Society 21st annual meeting. Santa Monica, California: Human Factors Society, 1977.



Appendix B

Beyond Averaging II: Single-Trail Classification of  
of Exogenous Event-Related Potentials Using Stepwise Discriminant Analysis

Richard L. Horst and Emanuel Donchin

Electroencephalography and Clinical Neurophysiology, December, 1979

# Abstract

Stepwise discriminant analysis (SWDA) was used to determine whether single-trial event-related potentials (ERPs) were elicited by a checkerboard presented to the upper or lower visual half-field. Discriminant functions were computed on the basis of "training sets" constructed of upper and lower half-field ERPs, and applied to "test sets" of other ERPs elicited by the same stimuli. Individual-subject discriminant functions for data recorded at Pz classified the single ERPs in the test sets with a mean accuracy of 83.7% correct. The mean accuracy attained by individual-subject functions from the most discriminable scalp site for each subject was 87.8% correct, and that attained by an across-subjects function was 78.1% correct. Averaged ERPs showed the previously reported polarity reversal of corresponding exogenous components in the upper and lower half-field waveforms. Moreover, the SWDA procedure chose ERP time points at the latencies of these exogenous components for discriminating between the half-field ERPs. The results demonstrate that SWDA can accurately classify single ERPs in which the systematic variance is localized in exogenous components, having periods within the range of frequencies which typically comprise the background EEG.

## INTRODUCTION

While signal averaging is quite useful for extracting brain event-related potentials (ERPs) from the "noise" of the ongoing EEG, it has serious drawbacks. One major drawback is that signal averaging requires that stimuli be repetitively presented under the same conditions. Unfortunately, it is often difficult to maintain the subject, and the environment, under stable conditions for the period of time needed to present a sufficient number of stimuli. Many investigators have, therefore, attempted to extract information about ERPs from single trials (Zerlin and Davis, 1967; Donchin, 1969a; Ruchkin, 1971; Weinberg and Cooper, 1972; Bartlett et al., 1975; Schwartz et al., 1976; Squires, et al., 1976; Kutas et al., 1977; Coppola et al., 1978; Ruchkin and Sutton, 1978 a,b; Aunon and McGillem, 1979; see overview by John, et al., 1978).

Whether or not such single-trial techniques can be utilized in a particular situation depends on the experimental question under consideration. Clearly, signal averaging remains the method of choice when an accurate estimate of an unknown ERP waveform is required. But, as pointed out by Donchin (1969b), if certain constraints are satisfied, the application of other statistical techniques becomes appropriate. For example, it can sometimes be assumed that the ERP elicited by an experimental stimulus is one of a specified class of waveforms. The experimenter in such cases is interested in determining which of these possible waveforms has been elicited on each trial. The task then is to classify, rather than to estimate, the observed waveforms. Stepwise discriminant analysis (SWDA) provides, in such situations, a procedure for developing the needed classification rule. This rule, the "discriminant function," is computed to discriminate between groups of ERPs in a "training set." The group membership of the ERPs in the training set is assumed on some a priori basis. The discriminant function can then be used to classify a "test set" of ERPs whose group membership is "unknown."

Squires and Donchin (1976) have shown that SWDA can be used to classify single-trial ERPs with a high degree of accuracy. They computed a discriminant function from a training set of ERPs elicited by task-relevant, rare and frequent auditory stimuli presented in a Bernoulli series. This function, when applied to a test set of ERPs recorded from different subjects in a different experiment, classified correctly 81% of the trials as to whether they had been elicited by rare or frequent stimuli. Much of the systematic variance in these ERPs was concentrated in the P300 component. This endogenous ERP component, which is sensitive to cognitive variables, has a long peak latency (300 msec. or more), a large amplitude (often 10-25 uv.) , and a long period (typically 200-500 msec.). Thus it was possible to enhance the signal to noise ratio of P300 relative to background EEG by digitally filtering the single ERPs prior to applying SWDA. This filtering reduced the contribution of the 8-13 c/sec alpha rhythm and higher frequencies in the EEG. How well SWDA can classify waveforms when such low-pass filtering is improper remains to be determined.

We demonstrate here that SWDA can classify, with remarkable accuracy, ERPs in which the systematic variance is found in exogenous components of the ERP. These exogenous components, which are primarily affected by the physical characteristics of the stimuli, are typically of smaller amplitude than P300 and are of a period (usually 25-100 msec for the "middle latency" components, occurring 50-200 msec after stimulus onset) which puts them within the range of frequencies typically found in the background EEG during task performance (e.g. Thompson and Obrist, 1964; Walter et al., 1967; Legewie et al., 1969).

The data described in the present report were obtained in an experiment designed to determine the extent to which exogenous visual ERP components are affected by task variables which influence P300. This aspect of the study will be reported elsewhere (Horst and Donchin, in preparation). The study provided, in addition, a suitable data-base that could be used for evaluating the ability

of SWDA to discriminate and classify ERPs on the basis of differences in the exogenous components.

Striking differences in the visual ERP can be obtained by presenting a patterned stimulus to different retinal loci. For example, a patterned stimulus presented to the upper visual half-field elicits an ERP which, over an epoch ranging from about 50-150 msec after stimulus onset, is largely inverted in polarity relative to the ERP elicited by the same stimulus presented to the lower visual half-field. This effect has been reported by numerous investigators, using several modes of pattern stimulation (Jeffreys, 1971; Michael and Halliday, 1971; Jeffreys and Axford, 1972b; Lesevre, 1973; Purves and Low, 1976; Lehmann et al., 1977). Using a "pattern appearance-disappearance" stimulus, Jeffreys and colleagues (Jeffreys and Axford, 1972a,b; Jeffreys, 1977) reported that three distinct ERP components can be observed during this epoch. Based both on differences in the scalp distributions of these components, with reference to the well-established retinotopic mapping of visual cortex, and on the dissimilar ways in which they varied with the retinal locus of the stimulus, Jeffreys argued that these components, which he labelled CI, CII, and CIII, are generated in distinct loci in striate and extra-striate cortex.

Our approach was to use SWDA to discriminate and classify such half-field ERPs. Experimental blocks in which we replicated Jeffrey's conditions provided the training sets for SWDA. As test sets we used data from experimental blocks in which the subjects were required to count occurrences of various subsets of the stimuli. To the extent that SWDA accurately classifies single-trial ERPs as to whether they were elicited by upper or lower half-field stimuli, we demonstrate the utility of the technique for discriminating ERPs which vary in their exogenous components, and confirm the trial to trial reliability of the half-field effects seen in ERP averages.

## METHOD

### Subjects

Young adult volunteers, four males and six females, were paid for their participation in the experiment. Subjects were tested in a single, 2 hour experimental session. All subjects had normal or corrected-to-normal vision. None of the subjects had previously participated in an ERP experiment. The purpose of the experiment was explained at some length and subjects were repeatedly exhorted during data acquisition to maintain eye fixation. One additional subject reported a persistent double image of the fixation cross, so her session was terminated and the data discarded.

### Stimuli and Apparatus

Subjects semi-reclined in an easy chair in a dimly lighted room. They viewed high contrast, lithographic negatives back-illuminated in a three-channel tachistoscope. The field in each channel subtended 6 degrees by 6 degrees of visual angle. One field was covered by a 50% transmittance neutral density filter and contained a centrally located, opaque fixation cross (horizontal and vertical bars both about 1 degree in extent and about 1 minute in thickness). The other two fields contained, in one half, a 3 degree by 6 degree checkerboard pattern consisting of transparent and opaque checks 20' on a side, and in the other half, a 50% transmittance neutral density filter. In one channel the checkerboard was positioned in the upper half of the field; in the other channel it was positioned in the lower half of the field. The medial horizontal border of each checkerboard half-field was aligned to be contiguous with the horizontal bar of the fixation cross in the third channel. The checkerboard fields were illuminated separately for either a short (25 msec) or a long (125 msec) exposure duration. Thus there were four possible checkerboard stimuli -- upper-short, upper-long, lower-short, and lower-long. The fixation field was constantly

illuminated except during the presentation of one of these stimuli. Thus the appearance and disappearance of the checkerboard stimuli involved no change in luminance of the unpatterned half-field and no change in mean luminance of the patterned half-field. The constant luminance was 25 footlamberts. Stimulus presentation and on-line monitoring of data collection were controlled by a mini-computer.

### Recording

EEG was recorded from seven midline scalp sites, each referred to the linked earlobes. Subjects were grounded either at the chin or the forearm. The scalp sites were 10-20 system locations Fz, Cz, Pz, and Oz, plus sites midway between Cz and Pz, midway between Pz and Oz, and at theinion. Burden Ag-AgCl electrodes were affixed to these scalp sites with collodion. EOG was recorded between sites inferior-lateral to the one eye and superior lateral to the other eye. Beckman Biopotential electrodes were affixed to the face, earlobes, and in some cases, forearm with adhesive collars. Electrode impedances were always less than 10 kohms, and usually less than 5 kohms. The amplifiers had an upper half-amplitude of 60 Hz, with a 60 Hz notch filter, and a time constant of .8 sec. EEG and EOG were recorded, along with event markers, on analog tape. Single trial ERPs were digitized off-line and stored on digital tape. EEG was digitized every 4 msec for an epoch extending from 100 msec before to 800 msec after the presentation of each checkerboard stimulus. The analyses reported here concern the first 400 msec of this epoch.

### Procedure

The four stimuli were presented in random-appearing sequences and at inter-stimulus intervals which varied unpredictably, in 10 msec steps, between 1400 and 1600 msec. There were twenty-four blocks of trials. In twenty of these blocks each of the four stimuli was presented 25% of the time. In blocks 3, 4,

21, and 22 the stimuli were not equiprobable, but data from these blocks were not included in the present analyses.

The variance of the EOG for each stimulus presentation was calculated on-line. If this variance exceeded a value chosen during pilot work to detect blinks and eye movements of 2 degrees or more, the ERP was rejected and that stimulus was reinserted in the sequence 3 to 5 trials later. Thus each block resulted in 100 "good EOG" trials [1].

The task required of subjects varied across blocks. In order to replicate the conditions imposed in most of the previously reported work with checkerboard stimuli, we instructed subjects during blocks 1 and 2 simply to focus on the center of the cross and "attend" to the stimuli. They were told to blink if they had to, but then to immediately return to focusing on the cross. After block 2, the counting task was mentioned for the first time. Subjects were instructed to keep a covert count of the number of occurrences of a designated pair of stimuli, and to avoid movements of the head, mouth, tongue, or extremities. There were four different counting tasks -- to count all upper half-field stimuli regardless of exposure duration, to count all lower half-field stimuli regardless of exposure duration, to count all short exposure duration stimuli regardless of half-field, and to count all long exposure duration stimuli regardless of half-field. These four tasks were counterbalanced across blocks 5 through 20 in a Latin-square design. Before each block, subjects were informed of the counting task to be performed and of the stimulus probabilities, and were reminded to focus on the cross. The subjects' counts were monitored to ensure an acceptable level of performance. In blocks 23 and 24, subjects were again requested to fixate the cross and "attend" to the stimuli, but were explicitly told not to count. Blocks 1, 2, 23, and 24 therefore constituted the "no task" condition. There was a one minute pause between blocks and, midway through the session, a 15 minute break occurred during which subjects were disconnected from the EEG



amplifiers and allowed to move about.

#### Data analysis

Average ERPs were constructed for all combinations of subjects X tasks X stimuli X scalp sites, averaged over the four appropriate blocks for each task. Detailed analyses of these averages will be presented elsewhere (Horst and Donchin, in preparation). Of interest here are the average differences between the groups of single ERPs which were submitted to SWDA. Principal components analysis was used to quantify the differences among these average ERPs.

Single trial ERPs, consisting of the 75 digitized voltages between stimulus onset and 300 msec post-stimulus, were subjected to SWDA (BMD07M, see Dixon, 1970). As training sets for the development of discriminant functions, we used the ERPs elicited by the two shorter duration stimuli presented during the four "no task" blocks. Thus discriminant functions were computed to distinguish the ERPs elicited by upper-short stimuli from those elicited by lower-short stimuli. Separate "individual-subject" discriminant functions were computed on data from each subject at each of three scalp sites -- Cz, Pz, and Oz. In addition, an "across-subjects" discriminant function was computed using the Pz data of all subjects. These discriminant functions were first used to classify the single ERPs on which they were derived, i.e., the upper-short and lower-short ERPs from the "no task" blocks. Then each subject's discriminant function derived from Pz data was used to classify a test set consisting of single-trial ERPs recorded at Pz from that subject in the counting task blocks (5-20). For some subjects, as described below, the discriminant function based on "no task" data from Oz was also applied to Oz data from the counting task blocks. Finally, the across-subjects discriminant function was applied to each subject's Pz ERPs from the counting task blocks.

## RESULTS

The extent to which we replicate the results reported previously by Jeffreys and his associates can be judged from the average ERPs shown in Figure 1. Here ERPs elicited by the upper-short and lower-short stimuli during the "no task" blocks are superimposed. Data are shown for each of three scalp sites (Cz, Pz,

-----  
Insert Figure 1 About Here  
-----

and Oz) for each subject. The expected polarity reversal in the first two and sometimes three peaks of the waveforms is evident in eight of the ten subjects (for subject BD the lower half-field ERP is not well-defined; for subject LH there is considerable alpha activity remaining in the averages). Due to individual differences in the scalp distributions of the various peaks, the scalp site of the maximal difference between upper and lower half-field ERPs varied somewhat across subjects.

Principal-components analysis of average ERPs

These differences were quantified by a principal-components analysis (PCA) (see Donchin, 1966; Squires et al., 1977; Donchin and Heffley, in press). The data set for PCA consisted of average ERPs from all scalp sites and all subjects for the upper-short and lower-short stimuli in the "no task" condition. The cross-products matrix of association among the variables (ERP time points) was factored, and the seven components which accounted for the largest percentages of the total variance among the waveforms in the data set were varimax rotated [2]. Component loadings for the first four components extracted are plotted in Figure 2A. These components accounted for respectively, 42.6, 21.3, 15.7, and 9.3% of the total variance among the waveforms in the data set. These loadings represent

-----  
 Insert Figure 2 About Here  
 -----

ERP regions which varied orthogonally. To see if the variance represented by these component loadings was systematically related to the experimental variables, an analysis of variance (10 subjects with repeated measures on 2 half-field stimuli x 7 scalp sites) was performed on the component scores for each component. These component scores are measures of the extent to which each component loading is represented in each ERP (see Donchin and Heffley, in press, for the manner in which these scores are computed). Component 1 did not vary systematically with either half-field or scalp site, and probably reflects differences among the subjects. For component 2, both the effect of stimuli ( $F=15.10$ ;  $df=1,9$ ;  $p<.01$ ) and the interaction between stimuli and scalp sites ( $F=6.22$ ;  $df=6,54$ ;  $p<.001$ ) were statistically significant. Similarly, component 3 varied significantly with stimuli ( $F=19.96$ ;  $df=1,9$ ;  $p<.01$ ), scalp sites ( $F=7.51$ ;  $df=6,54$ ;  $p<.001$ ) and the interaction between stimuli and scalp sites ( $F=7.43$ ;  $df=6,54$ ;  $p<.001$ ). Component 4 was systematically related to stimuli ( $F=10.85$ ;  $df=1,9$ ;  $p<.01$ ) and scalp sites ( $F=3.32$ ,  $df=6,54$ ;  $p<.01$ ). The mean component scores for components 2, 3 and 4 are plotted in Figure 3. Based on their latencies, polarities, and scalp distributions, these three components can be identified with Jeffreys' components CIII, CI, and CII respectively. Moreover,

-----  
 Insert Figure 3 About Here  
 -----

the component scores for individual subjects reflected the individual differences in the scalp distributions of CI, CII, and CIII which are apparent in Figure 1.

It was necessary to determine if single ERPs from the counting task blocks

provided an appropriate test set for evaluating the discriminant functions derived from "no task" data. That is, did the stimuli presented in the counting task blocks elicit ERP components of the same latency, scalp distribution, and approximate amplitude as those presented in the "no task" blocks? To address this question, a PCA was performed on the upper-short and lower-short average ERPs from the counting task blocks. After a varimax rotation, the four components which accounted for the most variance in this PCA loaded at practically identical regions of the ERP epoch as did those in Figure 2. This finding suggests that the task demands of counting did not alter the latencies of ERP components CI, CII, and CIII. It was appropriate, therefore, to enter average ERPs from both the "no task" and the counting task blocks into a PCA to test for amplitude or scalp distribution differences between the ERPs elicited in these two conditions. Again the four components which accounted for the most variance showed, after a varimax rotation, component loadings similar to those in Figure 2. For each component, subject, scalp site, and half-field stimulus the component scores for the four counting tasks were averaged together for comparison with the "no task" component scores. An analysis of variance (10 subjects with repeated measures on 2 conditions -- "no task" vs. mean of the counting tasks -- X 2 half-field stimuli X 7 scalp sites) was performed on the component scores for each component. Statistically significant trends similar to those found previously (see Figure 3) again emerged. Moreover, there were no significant differences associated with task conditions. Thus the imposition of the counting tasks did not, at least on the average, alter the amplitudes or scalp distributions of ERP components CI, CII, and CIII. Therefore, ERPs from the counting tasks seem to be appropriate test data for the discriminant functions built on "no task" data.

#### Stepwise discriminant analysis of single-trials

\* detailed exposition of the use of SWDA can be found elsewhere (Donchin and

Herning, 1975; Donchin and Heffley, in press). In essence, SWDA identifies the time points (latencies) along the ERP epoch which best distinguish between the groups of ERPs in the training set. The discriminant function is a linear combination of ERP voltages at these selected time points (see Donchin and Heffley, in press). For each of the discriminant functions computed here, SWDA formed a weighted combination of the six time points which best discriminated between the single ERPs elicited by upper-short and lower-short stimuli in the "no task" blocks. The choice of the number of time points to be used in the function is somewhat arbitrary. A simulation conducted by Donchin and Herning (1975) suggested that little improvement is introduced by adding more than six points. Furthermore, examination of our obtained values of  $U$ , a statistic which provides an index of the separation between the two groups of ERPs at each step of the analysis (see Donchin and Herning, 1975), indicated that for most of the present analyses, discriminability increased only slightly beyond the first two or three time points selected.

Since the PCA of "no task" ERPs indicated that systematic differences between upper-short and lower-short averages occurred in ERP components CI, CII, and CIII, it is of interest to see whether SWDA selected time points from these components as the basis for distinguishing the two groups of single-trials. Table I presents the ERP time points, in the order in which they were selected, for each of the individual-subject discriminant functions. Figure 2B shows a

-----  
 Insert Table I About Here  
 -----

histogram of these time points, summed across subjects, from the analyses of Pz ERPs. Similar histograms were obtained from the analyses of Cz and Oz data. The

across-subjects discriminant function was:

$$Y = - .008 X_{116} + .007 X_{92} + .003 X_{236} - .003 X_{64} + .003 X_{160} - .002 X_{192} - .171$$

where the X variables represent, in order, the time points chosen. It is evident that all of the SWDAs selected time points primarily from the ERP regions in which components CI and CII are found.

As a preliminary indication of the ability of SWDA to reliably discriminate upper-short from lower-short ERPs, the SWDA program classified the trials in the training sets (see Dixon, 1970). Correct classification of an ERP was assumed to occur when that ERP was assigned to the class (upper-short or lower-short) associated with the stimulus that elicited it. Table II shows the accuracy with which the ERPs in the training sets, for each subject and scalp site, were classified. Since across subjects at a given scalp site there were no

-----  
Insert Table II About Here  
-----

systematic differences in the accuracy with which lower-short and upper-short ERPs were classified, each entry in Table II is the mean accuracy with which the two stimuli were classified by a particular discriminant function. In general, higher classification accuracies occurred at scalp sites where the separation between the average ERPs for upper and lower half-fields was greatest (see Figure 1).

For some subjects the most accurate classification occurred at Pz. The second most accurate classification for some of these subjects occurred at Oz, but for some it occurred at Cz. For other subjects, classification was most accurate at Oz. The second most accurate classification for all of these subjects occurred at Pz. Thus Pz seems to be the best scalp site for making comparisons across subjects. The mean accuracy with which the individual-subject

Pz discriminant functions correctly classified the training sets was 89.4%, ranging across subjects from 75.4 to 99.5%.

A necessary test of whether a discriminant function has identified systematic differences between groups is to attempt to classify sets of data other than those which were used to derive the function (see Lachin and Schacter, 1974). The single ERPs elicited by upper-short and lower-short stimuli in the counting task blocks provided such a test set. There were 800 of these ERPs in each subject's test set for a particular scalp site. The classification of each test set ERP was accomplished by multiplying the coefficients in the discriminant function by the voltages in that ERP at the selected time points and summing them, with the function's residual term, to form a "discriminant score." ERPs with a discriminant score greater than or equal to zero were classified as upper-shorts; those with a discriminant score less than zero were classified as lower-shorts. That this criterion was a reasonable one is indicated in Figure 4, which shows the distribution of discriminant scores for each stimulus from the application of the individual-subject Pz discriminant functions to test set data.

-----  
 Insert Figure 4 About Here  
 -----

The percentages of correct classification which resulted from the application of the various discriminant functions to the test sets are shown in Table III. Since there were again no significant differences between the

-----  
 Insert Table III About Here  
 -----

accuracy with which upper-short and lower-short ERPs were classified, the means of the two classification accuracies are presented. The individual-subject Pz

functions classified Pz test sets from the same subjects with a mean accuracy of 83.7% correct, ranging across subjects from 64.7-95.9% correct [3]. The subjects for whom accuracy was the lowest were the same subjects who had shown the lowest accuracy of classification of the training sets.

Even better classification was obtained by focusing on different scalp sites for different subjects. When for each subject we took the discriminant function from the scalp site which best classified that subject's training set (see Table II), and applied that function to test set data from the same subject and scalp site, mean classification accuracy rose to 87.8% correct, ranging from 76.8-95.9% across subjects (see Table III).

Furthermore, we asked how much loss in accuracy occurs with the gain in generality offered by the across-subjects discriminant function. Pz ERPs in the test sets were classified correctly by the across-subjects function at a still impressive average of 78.1%, ranging across subjects from 62.4 to 90.7% (see Table III). This high level of classification accuracy attests to the similarity among subjects of the latencies at which systematic differences between upper-short and lower-short ERPs occurred (see Table I and Figure 1). Analyses of variance indicated that for neither stimulus were the discriminant scores systematically related to the four different counting tasks.

Finally, it is of interest to examine the ERPs which were misclassified. Figure 5 shows, for each stimulus, grand averages across subjects of the ERPs in the test sets which were correctly and incorrectly classified by the individual-subject Pz functions. It is evident that on the average the

-----  
 Insert Figure 5 About Here  
 -----

misclassified ERPs have a shape similar to the ERPs for which they were mistaken. This trend was apparent in the average waveforms from each subject. The



waveshapes which resulted in misclassification may have occurred fortuitously, due to random fluctuations in the EEG; or they may have been, to some extent, systematic. It is possible that on some of these misclassified trials subjects failed to direct their gaze at the fixation cross, and may have instead viewed the field in such a way that the wrong retinal half-field was stimulated. Such trials would then have been classified correctly with respect to the retinal half-field actually stimulated, and our percentages of correct classification would be underestimates of SWDA's accuracy. An independent trial by trial measure of subjects' locus of eye fixation would be necessary in order to identify such trials.

#### DISCUSSION

The remarkable accuracy with which half-field ERPs were classified here strikingly demonstrates the power of SWDA for single-trial ERP analyses. Discriminant functions built on one set of single-trial data were successfully applied to other sets of single ERPs from the same subjects, stimuli, and scalp sites, but from different experimental blocks. It is this ability of SWDA to generalize to new data sets which makes the technique attractive for clinical diagnosis and for on-line monitoring of the performance of human operators. Classification accuracies were increased (overall mean, 87.8% correct) when, as might be feasible for recording repeatedly from the same operator, we used an individualized discriminant function from the scalp site which, for each subject, best classified training set data. Still quite high classification accuracies (overall mean, 78.1% correct) were obtained when, as might be necessary in a clinical screening procedure, we used an across-subjects discriminant function at a single scalp site.

These high classification accuracies are similar to those previously attained when classifying single rare and frequent auditory ERPs with SWDA

(Squires and Donchin, 1976). On the one hand, the present degree of success might have been predicted on the basis of the differences between average half-field ERPs. Because of the polarity reversal between corresponding components in upper and lower half-field ERPs, the absolute voltage difference between these average ERPs approached the difference in P300 typically found between the average ERPs to rare and frequent, task relevant stimuli (see e.g. Squires and Donchin, 1976; Duncan-Johnson and Donchin, 1977). On the other hand, however, the exogenous ERP components which were chosen by SWDA to discriminate upper and lower half-field ERPs (Figures 2) have much shorter periods than P300, and are thus closer to frequencies which have been shown to comprise the background EEG during the performance of tasks comparable to ours (e.g. the visual discrimination conditions of Walter et al., 1967).

Several previous investigations have applied SWDA to discriminate single-trial ERPs which differed in their exogenous components (Donchin, et al., 1970; Donchin and Herning, 1975; Purves and Low, 1978). However, the degree of discriminability achieved in these studies, while sufficient to provide evidence for reliable differences between the groups to which SWDA was applied, did not approach the high levels attained when classifying ERPs which differed primarily in P300 amplitude (Squires and Donchin, 1976). It was not apparent whether these previous failures to classify exogenous ERPs with high accuracy reflected a basic limitation of SWDA, or merely the fact that the amplitude differences between the ERPs being discriminated were small compared to the difference in P300 amplitude between the ERPs to rare and frequent, task-relevant stimuli. An important implication of the present results is that if the differences in exogenous component amplitude between two groups of ERPs are sufficiently large, the frequency composition of these ERPs does not preclude SWDA from discriminating them with a high degree of accuracy.

In addition, our results offer strong support for the trial-to-trial

reliability of the effects of visual half-field on the ERP. The present average ERPs elicited by upper-short and lower-short stimuli replicated those reported by Jeffreys and Axford (1972b). As in their work, we had no provisions for independently measuring the subjects' locus of eye fixation. In lieu of such a measure, we must assume that the subjects were complying with our instructions to focus on the fixation cross. This assumption seems, to a large extent, vindicated by the results. The accuracy with which SWDA classified single ERPs as to the presumed half-field of the stimuli which elicited them, implies both a high degree of homogeneity among each group of half-field ERPs and consistent differences between the groups.

In a more practical vein, the present results suggest the possibility of using the effects of retinal locus on ERPs as a trial-to-trial index of the direction of gaze. Whereas, in the present situation we inferred the half-field of the external stimulus, assuming a known direction of gaze, the same technique might prove useful when it is desired to infer the direction of gaze (by inferring the retinal locus from the ERP), given a known location of the external stimulus. Such an index might be useful both as an experimental control, to monitor the extent to which subjects are complying with instructions to fixate a particular locus in space, and as one measure of the performance of human operators interacting with complex displays.

Vidal (1977) has demonstrated the feasibility of such an approach. His subjects were instructed to direct their gaze to one of four fixation points, one on each side of a display, depending upon which way they wanted a computer-controlled cursor to move. By flashing a checkerboard in the center of the display, and processing the elicited ERP with discriminant analysis techniques, Vidal obtained four-way classification accuracies as high as our two-way accuracies. However, since the emphasis in the Vidal study was on maximizing the performance of the "biocybernetic" system as a whole, SWDA was augmented in

several ways. Single ERPs were pre-processed with a Wiener filter, a default category was defined for trials too equivocal to classify (however, these trials were taken into account by Vidal's measure of mutual information), and subjects received feedback as to whether the system's classification had been successful. Thus Vidal's study did not assess the efficacy of SWDA, by itself, for dealing with exogenous ERPs. Our experimental conditions, at the same time better controlled but more artificial than those of Vidal, allowed such an assessment.

How small the difference in direction of gaze which can be measured with the present technique remains to be determined. Furthermore, other possible influences on these ERP components, such as those of selective attention (Van Voorhis and Hillyard, 1977) and accommodation (Harter and Salmon, 1971), need to be investigated.

Nonetheless, it should be emphasized that, at least for well-defined half-field stimuli, the present classification accuracies may not be the upper limit attainable. We have already mentioned the possibility that some apparently misclassified ERPs may have really been correct classifications from trials on which subjects misdirected their gaze. Furthermore, since there were individual differences in the scalp site which best discriminated upper and lower half-field ERPs, some site other than the three investigated here may prove to be optimal for certain subjects. Finally, the analyses of the present average ERPs are consistent with previous reports (e.g. Jeffreys and Axford, 1972b; Jeffreys, 1977) in showing that ERP components at corresponding latencies in the upper and lower half-field waveforms vary not only in their polarity, but also in their scalp distributions (see Figures 1 and 3). However, each of the discriminant functions constructed here took into account data from only one scalp site. If scalp distribution information could be incorporated into the discriminant functions, classification accuracies might be further enhanced.

## References

- Aunon, J. I., and McGillem, C. D. Detection and processing of individual components in the VEP. Psychophysiology, 1979, 16, 71-79.
- Bartlett, F., John, E. R., Shimokochi, M., and Kleinman, D. Electrophysiological signs of readout from memory. II. Computer classification of single evoked potential waveshapes. Behavioral Biology, 1975, 14, 409-449.
- Coppola, R., Tabor, R., and Buchsbaum, M. S. Signal to noise ratio and response variability measurements in single-trial evoked potentials. Electroencephalography and Clinical Neurophysiology, 1978, 44, 214-222.
- Dixon, W. J. BMDP-biomedical computer programs. Los Angeles: University of California Press, 1970.
- Donchin, E. A multivariate approach to the analysis of average evoked potentials. IEEE Transactions on Bio-Medical Engineering, 1966, BME-13, 131-139.
- Donchin, E. Discriminant analysis in average evoked response studies: The study of single trial data. Electroencephalography and Clinical Neurophysiology, 1969, 27, 311-314. (a)
- Donchin, E. Data analysis techniques in average evoked potential research. In E. Donchin & D. B. Lindsley (Eds.), Average evoked potentials: Methods, results and evaluations. Washington, D.C.: U.S. Government Printing Office, NASA SP-191, 1969, pages 199-236. (b)
- Donchin, E., Callaway III, E., and Jones, R. T. Auditory evoked potential variability in schizophrenia. II. The application of discriminant analysis. Electroencephalography and Clinical Neurophysiology, 1970, 29, 429-440.
- Donchin, E., and Heffley, E. Multivariate analysis of event-related potential data: A tutorial review. In D. Otto (Ed.), Multidisciplinary perspectives in event-related brain potential research. Washington, D.C., U.S. Government Printing Office, EPA-600/9-77-043, in press.
- Donchin, E., and Herning, R. I. A simulation study of the efficacy of stepwise discriminant analysis in the detection and comparison of event related potentials. Electroencephalography and Clinical Neurophysiology, 1975, 38, 51-68.
- Duncan-Johnson, C. C., and Donchin, E. On quantifying surprise: The variation in event-related potentials with subjective probability. Psychophysiology, 1977, 14, 456-467.
- Harter, M. R. and Salmon, L. E. Evoked cortical responses to patterned light flashes: Effects of ocular convergence and accommodation. Electroencephalography and Clinical Neurophysiology, 1971, 30, 527-533.
- Jeffreys, D. A. Cortical source locations of pattern-related visual evoked potentials recorded from the human scalp. Nature, 1971, 229, 502-504.
- Jeffreys, D. A. The physiological significance of pattern VEPs. In J. E. Desmedt (Ed.), Visual Evoked Potentials in Man: New Developments. London: Oxford

University Press, 1977.

Jeffreys, D. A. and Axford, J. G. Source locations of pattern-specific components of human visual evoked potentials I. Component of striate cortical origin. Experimental Brain Research, 1972, 16, 1-21. (a)

Jeffreys, D. A. and Axford, J. G. Source locations of pattern-specific components of human visual evoked potentials II. Component of extrastriate cortical origin. Experimental Brain Research, 1972, 16, 22-40. (b)

John, E. R., Ruchkin, D. S., and Vidal, J. J. Measurement of event-related potentials. In E. Callaway, P. Tueting, and S. H. Koslow (Eds.), Event-related brain potentials in man. New York: Academic Press, 1978.

Kutas, M., McCarthy, G., and Donchin, E. Augmenting mental chronometry: The P300 as a measure of stimulus evaluation time. Science, 1977, 197, 792-795.

Lachin, J. M. and Schacter, J. On stepwise discriminant analysis applied to physiologic data. Psychophysiology, 1974, 11, 703-709.

Lehmann, D., Meles, H. P., and Mir, Z. Average multichannel EEG potential fields evoked from upper and lower hemi-retina: Latency differences. Electroencephalography and Clinical Neurophysiology, 1977, 43, 725-731.

Legewie, H., Simonova, O., and Creutzfeldt, O. D. EEG changes during performance of various tasks under open- and closed- eye conditions. Electroencephalography and Clinical Neurophysiology, 1969, 27, 470-479.

Lesevre, N. Potentiels evoques par des patterns chez l'homme: influence de variables caracterisant le stimulus et sa position dans le champ visuel. In A. Fessard et G. Lelord (Eds.), Activites evoquees et leur conditionnement. INSERM, Paris, 1973: 1-22.

Michael, W. F., and Halliday, A. M. Differences between the occipital distribution of upper and lower field pattern-evoked responses in man. Brain Research, 1971, 32, 311-324.

Purves, S. J. and Low, M. D. Visual evoked potentials to a reversing-pattern light-emitting diode stimulator in normal subjects and patients with demyelinating disease. Electroencephalography and Clinical Neurophysiology, 1976, 41, 651-652.

Purves, S. J. and Low, M. D. Effects of stimulus shape on visual evoked potentials. The Canadian Journal of Neurological Sciences, 1978, 5, 313-319.

Ruchkin, D. S. Sorting of nonhomogeneous sets of evoked potentials. Communications in Behavioral Biology, 1971, 5, 383-396.

Ruchkin, D. S., and Sutton, S. Latency characteristics and trial-by-trial variation of emitted potentials. In J. E. Desmedt (Ed.), Cognitive components in cerebral event-related potentials and selective attention, Basel: Karger, 1978. (a)

Ruchkin, D. S. and Sutton, S. Emitted P300 potentials and temporal uncertainty. Electroencephalography and Clinical Neurophysiology, 1978, 45, 268-277. (b)

Schwartz, E. L., Ramos, A., and John, E. R. Cluster analysis of evoked potentials from behaving cats. Behavioral Biology, 1976, 17, 109-117.

Squires, K. C., and Donchin, E. Beyond averaging: The use of discriminant functions to recognize event related potentials elicited by single auditory stimuli. Electroencephalography and Clinical Neurophysiology, 1976, 41, 449-459.

Squires, K. C., Wickens, C., Squires, N. K., and Donchin, E. The effect of stimulus sequence on the waveform of the cortical event-related potential. Science, 1976, 193, 1142-1146.

Squires, K. C., Donchin, E., Herning, R. I., and McCarthy, G. On the influence of task relevance and stimulus probability on event-related potential components. Electroencephalography and Clinical Neurophysiology, 1977, 42, 1-14.

Thompson, L. W. and Obrist, W. D. EEG correlates of verbal learning and overlearning. Electroencephalography and Clinical Neurophysiology, 1964, 16, 332-342.

Van Voorhis, S., and Hillyard, S. A. Visual evoked potentials and selective attention to points in space. Perception and Psychophysics, 1977, 22, 54-62.

Vidal, J. J. Real-time detection of brain events in EEG. Proceedings of the IEEE, 1977, 65, 633-641.

Walter, D. O., Kado, R. T., Rhodes, J. M., and Adey, W. R. Electroencephalographic baselines in astronaut candidates estimated by computation and pattern recognition techniques. Aerospace Medicine, 1967, 38, 371-379.

Weinberg, H., and Cooper, R. The recognition index: A pattern recognition technique for noisy signals. Electroencephalography and Clinical Neurophysiology, 1972, 33, 608-613.

Zerlin, S. and Davis, H. The variability of single evoked vertex potentials in man. Electroencephalography and Clinical Neurophysiology, 1967, 23, 466-472.

## FOOTNOTES

- [1] The number of rejected trials per block varied considerably among subjects. The mean number rejected for individual subjects ranged from .5 to 13.1 trials.
- [2] The implications for ERP data of factoring the cross-products matrix, rather than the covariance or correlation matrices of association among ERP time points, are discussed by Donchin and Heffley (in press). Conceptually, an analysis of the covariance matrix, in that it extracts sources of variability with respect to the grand mean waveform, will extract as components only ERP peaks which vary within the data set. For most experimental questions, such an analysis is the one most appropriate. A PCA of the cross-products matrix, because it evaluates the variance with respect to a baseline, in addition extracts as components ERP peaks which are present but which do not vary among the waveforms in the data set. However, the polarity of the component scores derived from a PCA of the cross-products matrix corresponds to the actual polarity of the ERP components, with respect to baseline; whereas the polarity of components derived from a PCA of the covariance matrix reflects the direction of the components relative to the grand mean waveform. In the present data similar components were extracted by PCAs of both matrices; i.e. there were no reliable ERP peaks which were unaffected by the experimental variables. Therefore, to retain information about component polarity, we report here the PCA of the cross-products matrix.
- [3] It is worth noting that these individual-subject Pz discriminant functions, which were constructed to discriminate between the two shorter duration stimuli, classified the longer duration stimuli (upper-long, lower-long) from the counting task blocks as accurately as they classified the shorter duration stimuli from these blocks. This result is not surprising when it is considered that the discriminant functions selected time points primarily in the region of ERP components CI and CII (Table I and Figure 4) and that these components do not seem to be affected by stimulus duration (Jeffreys, 1977; Horst and Donchin, in preparation).



## Figure Legends

Figure 1.

Superimposed average ERPs (100 trials each) elicited by upper-short and lower-short stimuli, from the "no task" condition, for each subject at three scalp sites.

Figure 2a.

Component loadings for the first four components extracted by the PCA of "no task" ERPs.

Figure 2b.

Histogram of the latencies chosen, as best distinguishing upper-short from lower-short ERPs, by the individual-subject discriminant functions for Pz data. The six latencies chosen by the SWDA of each subject's data were summed across subjects.

Figure 3.

Mean component scores for the components identified as CI, CII, and CIII from the PCA of "no task" ERPs.

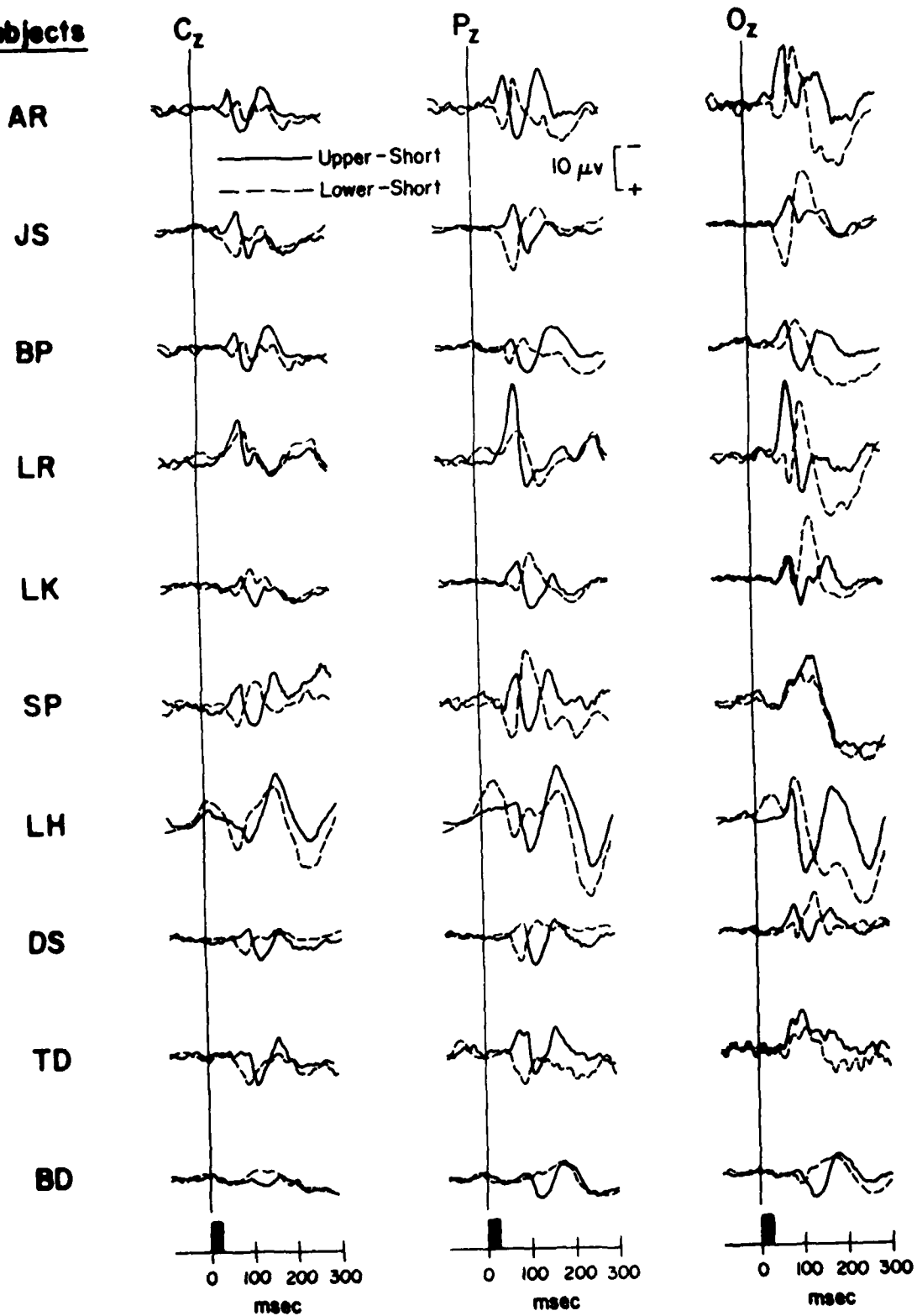
Figure 4.

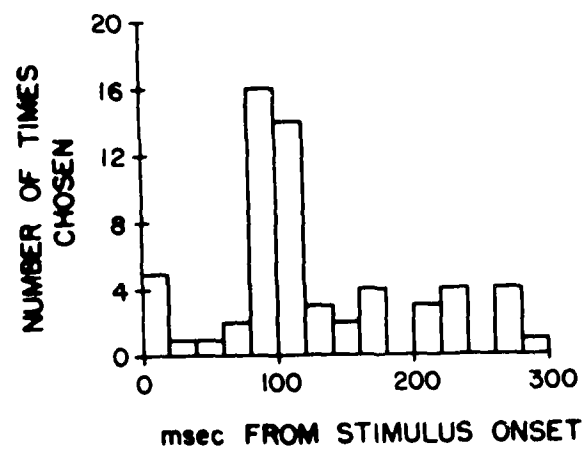
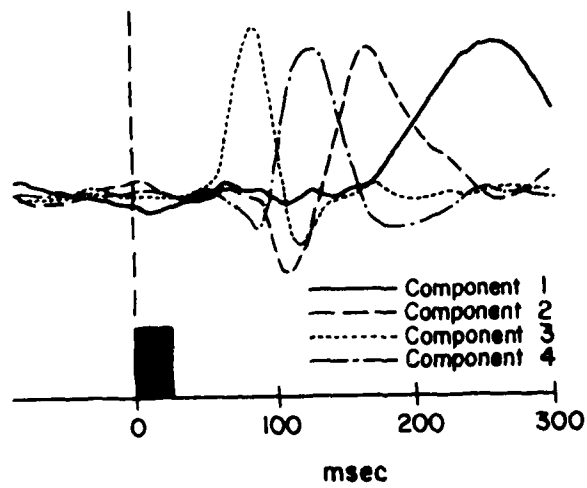
Histograms showing for each subject the distribution of discriminant scores, for ERPs elicited by upper-short and lower-short half-field stimuli, which resulted from the application of the individual-subject discriminant functions to Pz ERPs in the test sets. ERPs for which the discriminant score was less than zero were classified as lower-shorts; those for which the score was greater than or equal to zero were classified as upper-shorts.

Figure 5.

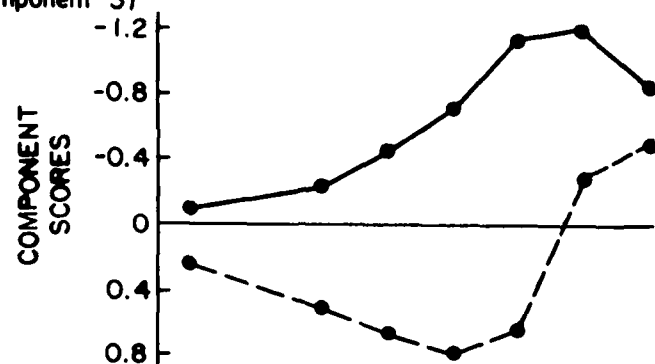
Grand averages over subjects of the test set ERPs, for each stimulus, which were correctly and incorrectly classified by the individual-subject discriminant functions for Pz data.

Subjects

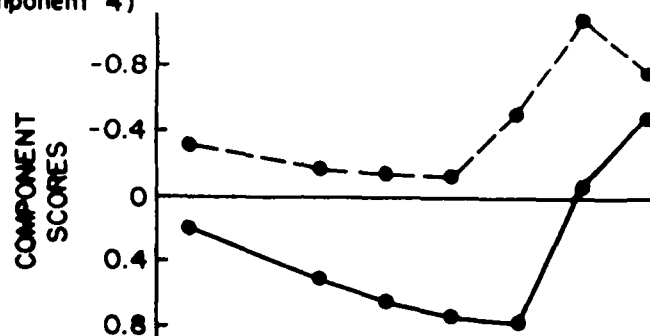




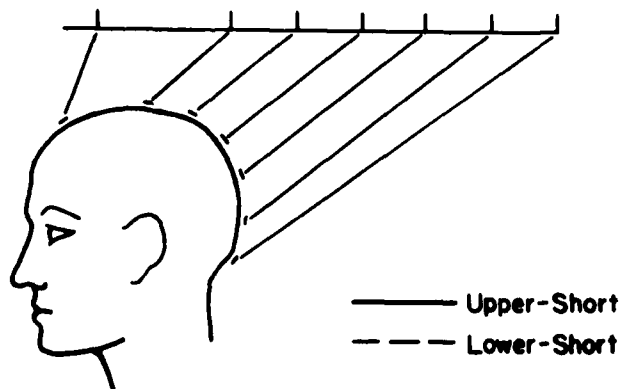
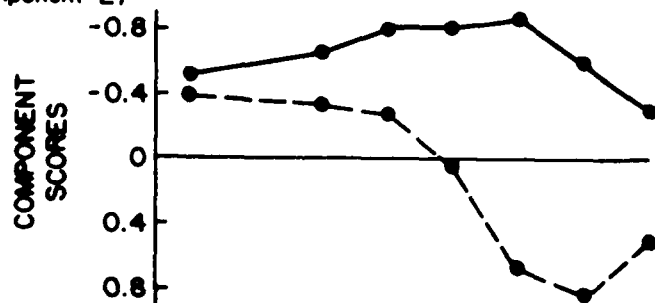
C I  
(Component 3)



C II  
(Component 4)



C III  
(Component 2)



Subjects

AR

JS

BP

LR

LK

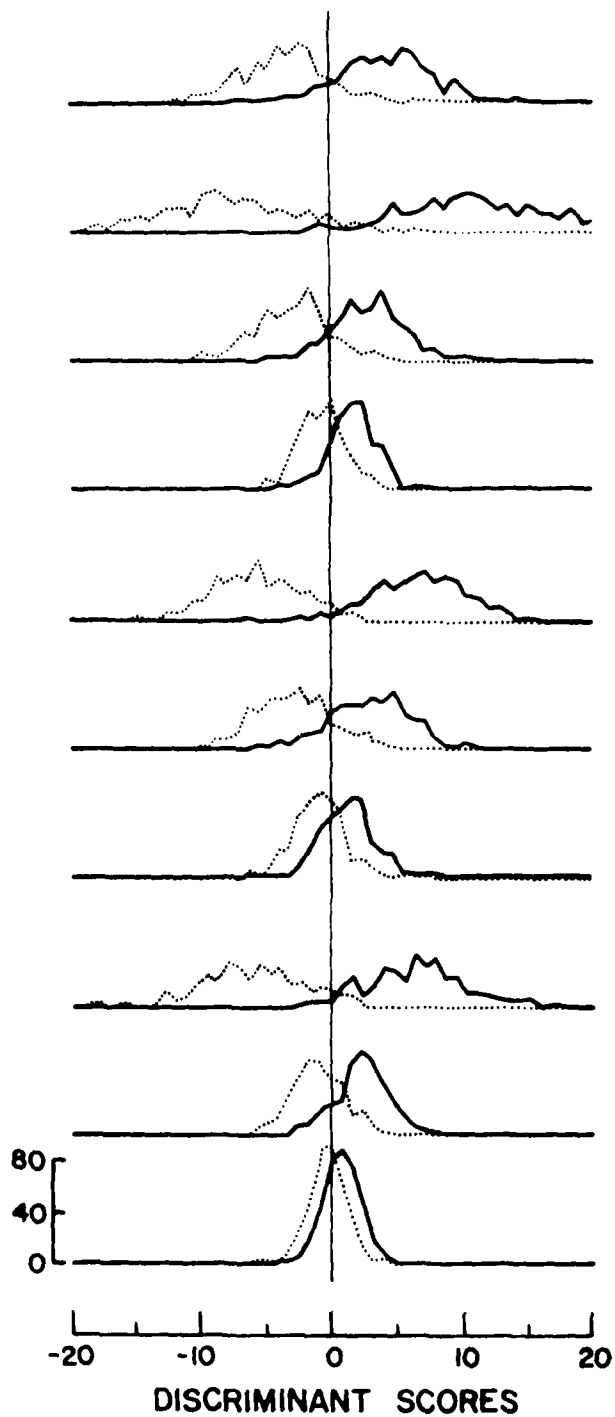
SP

LH

DS

TD

BD



— Upper Short

..... Lower Short

# SWDA Classification

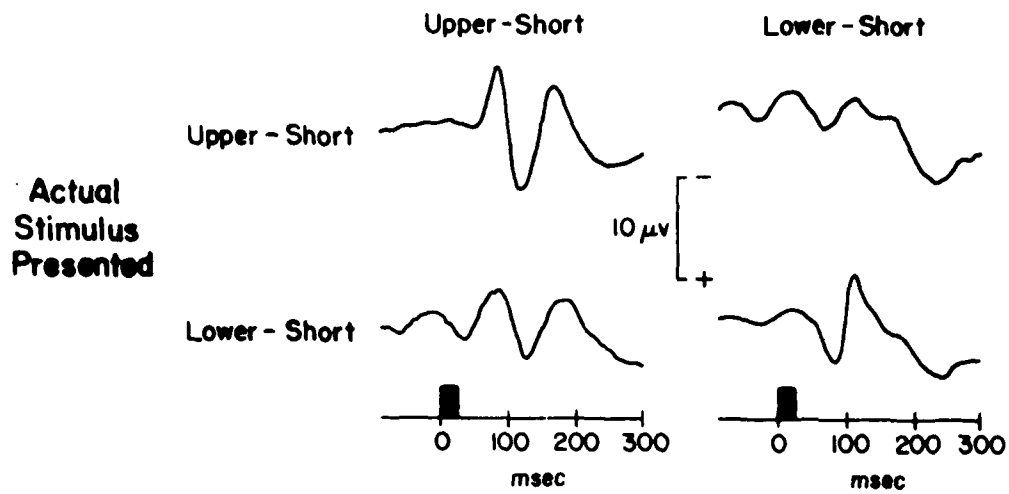


Table I

Latencies (msec from stimulus onset), in the order selected,  
which were chosen by the individual-subject discriminant functions  
as best discriminating upper from lower half-field ERPs

| <u>Subject</u> | <u>Scalp Site</u>           |                             |                              |
|----------------|-----------------------------|-----------------------------|------------------------------|
|                | <u>Cz</u>                   | <u>Pz</u>                   | <u>Oz</u>                    |
| AR             | 112, 88, 104, 160, 144, 128 | 108, 88, 208, 164, 276, 104 | 80, 112, 164, 88, 60, 188    |
| JS             | 96, 120, 72, 100, 176, 192  | 96, 120, 88, 4, 92, 160     | 96, 120, 100, 72, 4, 116     |
| BP             | 116, 92, 236, 64, 8, 88     | 116, 92, 228, 8, 64, 160    | 116, 192, 88, 72, 32, 56     |
| LR             | 116, 92, 4, 268, 204, 220   | 88, 112, 96, 4, 216, 268    | 120, 92, 220, 4, 128, 32     |
| LK             | 120, 92, 60, 172, 48, 188   | 120, 92, 176, 136, 112, 100 | 120, 176, 140, 96, 164, 116  |
| SP             | 88, 108, 164, 60, 196, 220  | 116, 232, 88, 108, 16, 64   | 124, 256, 220, 108, 164, 156 |
| LH             | 264, 108, 88, 64, 100, 232  | 232, 108, 36, 88, 120, 48   | 172, 228, 108, 48, 68, 116   |
| DS             | 92, 120, 48, 168, 184, 296  | 124, 88, 168, 116, 92, 296  | 124, 172, 92, 280, 240, 164  |
| TD             | 92, 116, 288, 160, 56, 236  | 92, 116, 232, 268, 164, 212 | 180, 160, 172, 272, 280, 284 |
| BD             | 116, 72, 172, 60, 4, 148    | 128, 4, 276, 96, 116, 88    | 128, 248, 92, 120, 16, 12    |

TABLE II

Percentages of training set trials correctly classified  
by the individual-subject discriminant functions

| <u>Subject</u> | <u>Scalp Site</u> |           |           |
|----------------|-------------------|-----------|-----------|
|                | <u>Cz</u>         | <u>Pz</u> | <u>Oz</u> |
| AR             | 79.1              | 94.9      | 92.4      |
| JS             | 87.0              | 99.5      | 99.5      |
| BP             | 75.7              | 87.9      | 95.2      |
| LR             | 70.3              | 81.9      | 97.8      |
| LK             | 83.9              | 94.3      | 99.0      |
| SP             | 86.5              | 90.6      | 69.8      |
| LH             | 77.4              | 85.2      | 87.2      |
| DS             | 84.2              | 95.9      | 95.4      |
| TD             | 79.8              | 87.9      | 73.5      |
| BD             | 64.0              | 75.4      | 89.1      |



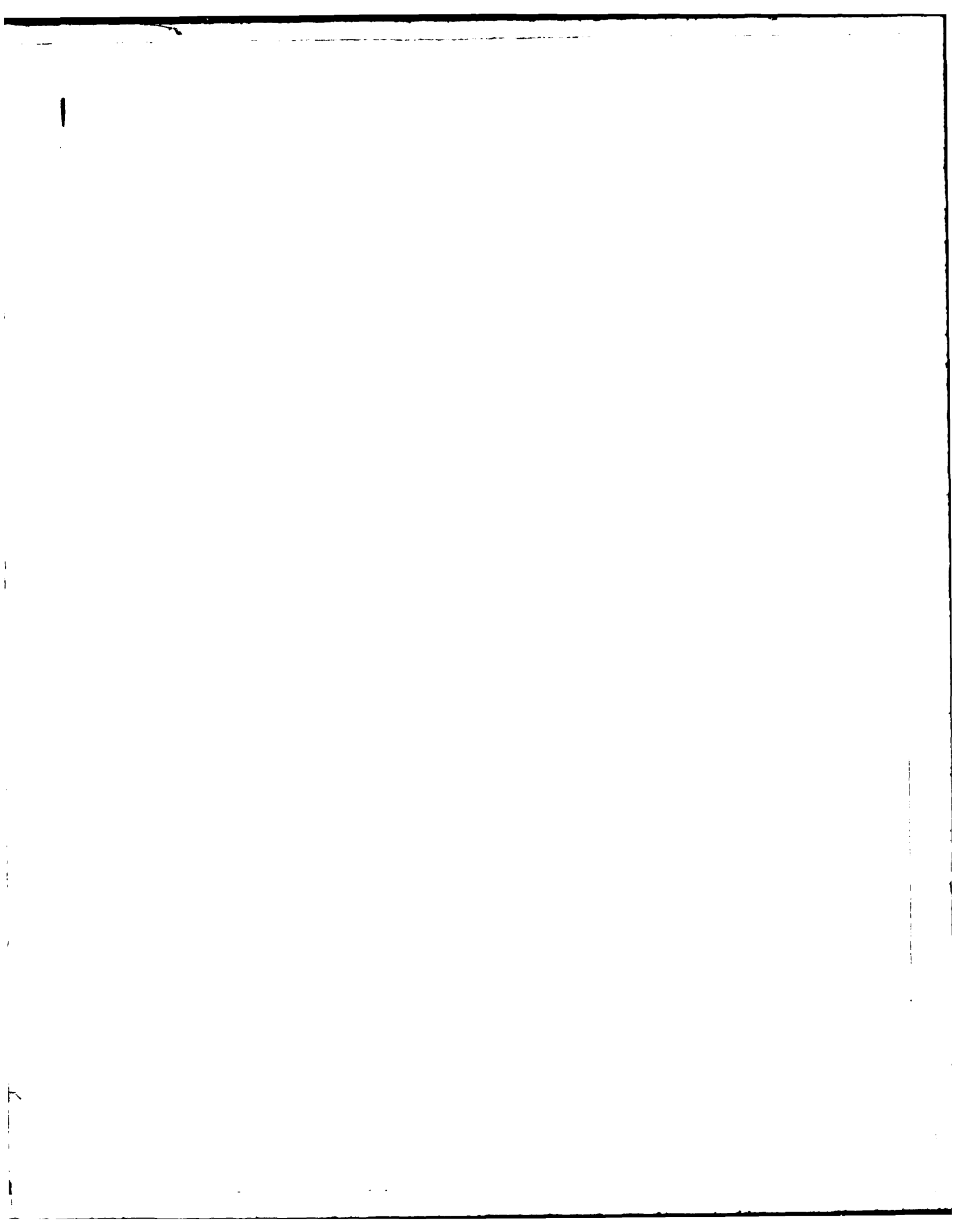


TABLE III

Percentages of test set trials correctly classified  
by the various discriminant functions

| <u>Subject</u> | <u>Individual-subject</u><br><u>Pz function</u> | <u>Individual-subject</u><br><u>Oz function*</u> | <u>Across-subjects</u><br><u>Pz function</u> |
|----------------|---|--|--|
| AR             | 88.1  |  | 80.8   |
| JS             | 95.7  |  | 85.4   |
| BP             | 85.4  | 90.7   | 84.3   |
| LR             | 77.6  | 95.7   | 66.2   |
| LK             | 95.9  | 95.9   | 90.7   |
| SP             | 84.0  |  | 79.8   |
| LH             | 72.2  | 76.8   | 62.4   |
| DS             | 93.7  |  | 85.2   |
| TD             | 79.5  |  | 81.9   |
| BD             | 64.7  | 78.5   | 64.2   |

\*The Oz discriminant functions were applied to test set data for only those subjects whose training set data had been classified more accurately at Oz than at Pz (see Table II).

Appendix C

Structural Interference in Dual Task Performance:  
Event-Related Potentials, Behavioral and Subjective Effects

Jack B. Isreal

Appendix D  
Event-Related Brain Potentials and Subjective  
Probability in a Learning Task

Richard L. Horst, Ray Johnson, Jr. and Emanuel Donchin  
Memory & Cognition (in press)

The  $P_{300}$ , one of several endogenous components in the human event-related potential (ERP) (see review by Donchin, Ritter, & McCallum, 1978), appears to be a manifestation at the scalp of brain activity associated with one or more cognitive processes. A  $P_{300}$  is elicited only by events which are both relevant to a task the subject is performing (Donchin & Cohen, 1967; Sutton, Tueting, Zubin, & John, 1967) and which resolve, for the subject, some uncertainty (Sutton, Braren, Zubin, & John, 1965). Furthermore, the latency of this component is proportional to the time it takes the subject to categorize the eliciting event (Kutas, McCarthy, & Donchin, 1977; Ritter, Simson, & Vaughan, 1972; N. Squires, Donchin, K. Squires, & Grossberg, 1977).

Much evidence supports the assertion that the amplitude of  $P_{300}$  varies inversely with the probability that the subject associates with the eliciting event. A completely predictable event, even if task-relevant, elicits little if any  $P_{300}$  (Donchin, Kubovy, Kutas, Johnson, & Herning, 1973; Friedman, Hakerem, Sutton, & Fleiss, 1973; Sutton et al., 1965). When there is uncertainty as to which of two stimuli will occur, the less frequently occurring event elicits the larger  $P_{300}$  (Sutton, et al., 1965). Moreover, when event probability is manipulated, systematic variations in  $P_{300}$  amplitude are obtained. Tueting, Sutton, and Zubin (1970), using a guessing task, were the first to show that as the prior probability of a stimulus was decreased, the amplitude of the elicited  $P_{300}$  increased (also see e.g., Friedman, et al., 1973; N. Squires, Donchin, Herning, & McCarthy, 1977). By parametrically varying stimulus probabilities in a counting task, Duncan-Johnson and Donchin (1977) demonstrated that  $P_{300}$  amplitude, for task relevant stimuli, was a decreasing function of prior probability over a range from .10 to .90.

In addition to the effect of prior probability,  $P_{300}$  amplitude varies with the sequence of preceding stimuli (see Tueting et al., 1970). At all levels of probability in the Duncan-Johnson and Donchin study a stimulus which had been preceded by itself elicited a smaller  $P_{300}$  than one which had been preceded by the other stimulus. Similarly, K. Squires, Wickens, N. Squires, and Donchin (1976) and K. Squires, Petuchowski, Wickens, and Donchin (1977) showed that the  $P_{300}$  elicited by a stimulus in a Bernoulli series is influenced by the sequence of stimuli presented on the preceding five trials. K. Squires et al. (1976) proposed that the subjective probability (or "expectancy") associated with a stimulus is a linear combination of the prior probability of that stimulus and the subject's exponentially decaying memory of the sequence of preceding stimuli. Assuming that  $P_{300}$  amplitude is inversely related to this subjective probability, their model accounted for 78% of the variance in  $P_{300}$  amplitude. This model is similar to models developed to account for sequential effects in choice reaction time (RT) (Audley, 1973; Falmagne, 1965; Laming, 1969).

The effects of event probability and sequence on  $P_{300}$  cannot be attributed to habituation or to receptor adaptation. For it appears to be the probability of stimulus categories, rather than the frequency with which particular physical stimuli occur, that governs the effects of both prior probability (E. Courchesne, Hillyard, & R. Courchesne, 1977; Friedman, Simson, Ritter, & Rapin, 1975; Kutas & Donchin, 1978; Tueting et al., 1970) and event sequence (Johnson & Donchin, in press) on  $P_{300}$ . Furthermore, the degree to which the previous sequence of stimuli affects the amplitude of  $P_{300}$  depends on task conditions. In a warned RT task, Duncan-Johnson and Donchin (1978) showed that sequential effects on  $P_{300}$  were eliminated when the warning stimulus provided information about the probability with which

particular imperative stimuli would occur.

In almost all previous studies in which the relation between subjective probability and  $P_{300}$  amplitude was examined, subjects derived the expectancies that they presumably assigned to events from attributes of the environment--the prior probabilities and the sequences in which the experimenter delivered stimuli. In the present study we attempted to determine if the relationship between  $P_{300}$  amplitude and subjective probability would hold when subjects formed expectancies on the basis of their changing knowledge about the environment. The subjects were assigned a classical paired-associate learning task. In response to the first ("stimulus") syllable of each pair, subjects typed the three-letter syllable which they thought was the paired-associate. They also reported their confidence in the correctness of this response. The correct paired ("response") syllable was then presented. The extent to which the correct "response" syllable was expected at this point was assumed to depend on the subjects' confidence in the correctness of their three-letter responses. As learning occurred, these internally formed expectancies should have changed even though there was no change in the manner with which external stimuli were being presented. Thus an analysis of the ERPs elicited by the "response" syllables according to the subjects' confidence ratings and to the trial outcomes (that is, whether or not their three-letter responses had in fact been correct), allowed an examination of the relationship between subjective probability, as inferred from subjects' own indications of their expectancies, and  $P_{300}$ .

K. Squires, Hillyard, and Lindsay (1973) have studied the amplitude of  $P_{300}$  elicited by stimuli which indicated to the subjects whether they had been correct or incorrect in the detection of a near-threshold auditory stimulus. They report that the amplitude of  $P_{300}$  was larger when the

feedback disconfirmed subjects' judgments. Whether the same effect would be obtained in a learning task, where the ERP-eliciting stimuli provided feedback as to the accuracy of associations being formed in memory, rather than the accuracy of a sensory discrimination, was of interest here.

## METHOD

### Subjects

Six students at the University of Illinois (three males) were paid for their participation in the experiment. Their ages ranged from 19 to 28. Four subjects had participated in previous ERP experiments. A seventh subject completed all three sessions, but his data were discarded because his confidence ratings were confined almost exclusively to the two extreme points of the rating scale.

### Apparatus and Stimuli

Subjects sat in an easy chair, positioned in front of a PLATO computer terminal (see Smith & Sherwood, 1976) and held a detachable keyboard in their laps. The ERP-eliciting stimuli were consonant-vowel-consonant (CVC) nonsense syllables presented on the plasma-panel display of the terminal (see Johnson, Bitzer, & Slottow, 1971). The CVCs subtended 0.6 deg by 0.2 deg of visual angle and were 3.2 fL in luminance, compared to the 0.2 fL background of the display. A continuously presented rectangle, that subtended 2.8 deg by 1.2 deg of visual angle, surrounded the area of the panel at which the CVCs appeared and served as a target for the subject's gaze. Ambient lighting was adjusted to a comfortable level for each subject.

The subjects learned from repeated presentations which "response" CVC was paired with each "stimulus" CVC. Lists of six paired CVCs were constructed with the following constraints: 1) all CVCs were of low meaningfulness (less than or equal to 1.50 on the  $m'$  scale--Noble, 1961),



2) no CVC appeared in more than one list, 3) the six "stimulus" CVCs were highly similar, usually differing from each other in only one or two letters, 4) the six "response" CVCs were much less similar--no two of them had the same consonant in a given position and no syllable contained any letters of the paired "stimulus" CVC.

The PLATO computer system controlled the presentation of stimuli and processed subjects' responses from the keyboard. A PDP 11/10 received synchronizing pulses and identifying information from the PLATO computer, digitized and processed the EEG, and allowed the experimenter to monitor data collection via a GT-40 display. Data analyses were performed, off-line, on a Harris /7 computer. The statistical packages SPSS (Nie, Hull, Jenkins, Steinbrenner, & Bent, 1975) and ALICE (Grubin, Bauer, & Walker, 1976) were used for data analysis.

#### Procedure for Paired-Associate Task

Events on each trial. As illustrated in Figure 1, after a 1000 msec foreperiod, during which the target rectangle was empty, a "stimulus" CVC was presented for 500 msec. Then, following a 1000 msec delay, three

-----  
Insert Figure 1 About Here  
-----

question marks were displayed in the rectangle, signalling the subject to respond. The subject then typed the three letters which he, or she, thought was the correct "response" CVC, followed by a confidence rating from 0 to 100. The subjects' responses were echoed on the PLATO display and appeared in the rectangle. The keystroke which terminated the confidence rating initiated a 1000 msec interval during which the rectangle was again empty. The correct "response" CVC was then presented for 500 msec. After a further delay of 1000 msec, three percent signs appeared in the rectangle,

signalling a four sec inter-trial interval (ITI). The offset of these percent signs initiated the next trial. If the subject struck any key before the three question marks appeared, failed to complete the responses within 15 sec, or entered an invalid confidence rating, three asterisks were displayed instead of the "response" CVC. ERPs were recorded both to presentations of the "stimulus" and "response" CVCs. In each case, the recording epoch extended for 1750 msec, starting 250 msec before CVC onset.

Learning paired-associate lists. The pair of syllables to be presented on each trial was selected at random from the five pairs in the list which had not been presented on the previous trial. This procedure was followed until the subject gave two consecutive correct responses to each of the six pairs in the list. If a subject, after twice responding correctly to a given "stimulus" CVC, subsequently responded to it incorrectly, two further correct responses were required. All subjects learned the same eight lists, two in the first session and three in both the second and third sessions, but in a randomized order.

Instructions. Before starting to learn each list, subjects were reminded to watch the target rectangle and, from the beginning of each CVC foreperiod until the question marks or percent signs appeared following a CVC, to avoid movements of the eyes, mouth, or body which could cause recording artifacts. The following instructions regarding the use of the confidence rating scale also appeared on the PLATO terminal prior to the presentation of each list:

We want to correlate your brain waves with your confidence ratings. So it is very important that on every trial you do the confidence rating as accurately as you can. Remember after entering a three-letter response you are to rate your confidence as to whether that response, as a whole, was correct or incorrect. The confidence scale is meant to represent a continuum of confidence from one extreme, where you are as sure as you can be that your response was incorrect (0--definitely incorrect), to the other extreme, where you are as sure as you can be that your response was correct (100--definitely correct).

As a general guideline, use a rating between 0 and 25 when you are

very sure that your three-letter response was incorrect; use a rating between 25 and 50 when you think your response was probably incorrect, but you are not so sure; use a rating between 50 and 75 when you think your response was probably correct, but you are not sure; use a rating between 75 and 100 when you are very sure that your response was correct.

Within these general guidelines, you should choose an integer which you feel reflects your confidence accurately, with relatively large numbers indicating more likely correct and relatively small numbers indicating more likely incorrect.

Remember that you should try to learn each list as fast as possible. If you have any questions, ask the experimenter now.

Determining confidence ranges. Pilot work, in which a four-point confidence rating was used, revealed marked individual differences in the manner with which subjects rate their confidence in the paired-associate task. Since the same numerical value appears to have different meanings to different subjects, it would be misleading to use the nominal values of the confidence ratings to classify the FRPs. In this study we used a 101-point confidence scale. This choice allowed us to partition each subject's scale, based on that subject's usage of the scale, into ranges that would be equivalent across subjects.

With the following procedure, each subject's data were partitioned into four such ranges of confidence. First, the 101-point scale was collapsed to a 21-point scale by combining the ratings in successive 5-point sections of the scale (rating 100 was treated as a "section" by itself). The ratings in these sections were then further grouped into "regions" of the scale (Figure 2a) that each contained 4% or more of all the ratings entered by that subject while learning all eight lists.<sup>1</sup> Next we determined the percentage

-----

Insert Figure 2 About Here

-----

of trials in each of these regions on which the subject entered the correct three-letter response (Figure 2b). Finally, with the constraint that only adjacent regions could be combined, the scale was further collapsed into

four "ranges" of confidence such that the combined trials best approximated 0, 33, 67, and 100 percent correct (Figure 2c).

This partitioning resulted in ranges of confidence that can be considered equivalent, in terms of percentage of correct trials, across subjects. Note that the partitioning was done only as a matter of convenience for examining averaged ERPs. No claim is made that the derived ranges correspond, in either number or boundaries, to confidence ranges which the subjects may have formed internally. Note further that since the partitioning was done without regard to the ERP data, we did not prejudge the existence of ERP differences among the four confidence ranges. For convenience, we will refer to the four ranges of confidence, those at which accuracy approximated 0, 33, 67, and 100%, as respectively the "certainly wrong," "probably wrong," "probably right," and "certainly right" ranges; however, we neither imply that the trials within a given range are homogeneous nor that the ranges necessarily represent symmetrical states of confidence.

#### Procedure for Counting Task

For comparison with the ERPs recorded in the paired associate task, we wished to obtain ERPs from our subjects while they performed a task in which a well-defined P300 is typically seen. Therefore, ERPs were also elicited by CVCs in a counting task. Lists of six single CVCs were constructed with the same constraints as the "response" CVCs of the paired-associate lists. No CVC appeared in both the count and paired-associate lists.

Each trial consisted of a 1000 msec foreperiod followed by the 500 msec presentation of a randomly selected CVC (other than the one which had just occurred). Then following a 1000 msec delay, three question marks appeared in the target rectangle, signalling a four sec ITI. With the disappearance of the question marks the foreperiod of the next CVC began. As in the

paired-associate task, ERPs and eye movements were recorded for 1750 msec, beginning 250 msec before CVC onset. A block of 60 counting task trials was presented at the beginning and end of each experimental session. Prior to each block, one of the six CVCs in the list was designated as the target and subjects were asked to keep a covert count of the number of times it occurred. At the end of the block, subjects typed their count (these were always accurate to within plus or minus one).

### Recording

EEG was recorded from frontal, central, parietal and occipital scalp sites (Fz, Cz, Pz, and Oz in the International 10-20 system) each referred to the linked mastoids. The electrooculogram (EOG) was recorded from sub- and supra-orbital sites, each referred to the linked mastoids. Subjects were grounded with a chin electrode. Burden Ag-AgCl electrodes, affixed with collodion, were used on the scalp. Beckman Biopotential electrodes affixed with adhesive collars were used for the EOG, ground and reference sites. Electrode impedances were always below 10 kohms. EEG and EOG were amplified by modified Grass model 7P122 amplifiers (with an upper half-amplitude of 35 Hz and a time constant of 8 sec). The PDP 11/10 sampled the EEG and EOG every 10 msec during the 1750 msec epochs. These digitized ERPs, along with identifying information, were written on magnetic tape.

### Analysis of ERPs

Trials with EOG activity sufficient to contaminate the scalp recordings were identified with a peak detection algorithm. Only trials free of contamination were included in the ERP analyses, whereas all trials were included in the analyses of behavioral data. Since variability in the latency of P300 among the paired-associate average ERPs made a principal-component analysis of the waveforms inappropriate (see Donchin &

Heffley, 1978), a base-to-peak amplitude measure of  $P300$  was employed. Since it was necessary to compare average ERPs that were composed of very different numbers of trials, average ERPs were first digitally low-pass filtered (half-power frequency--6.3 Hz, see Ruchkin & Glaser, 1978) to attenuate any high-frequency EEG activity that remained in the averages. Then the difference between the mean voltage of the pre-stimulus ERP points and the voltage of the most positive point between 350 and 950 msec after CVC onset was calculated.

## RESULTS

### Paired-Associate Behavioral Data

Trials to criterion. There was considerable variability both within and between subjects in the number of trials needed to learn a list. Across subjects, the mean number of trials to criterion was 56 (S.D. = 21). Repeated measures analyses of variance showed no systematic differences either within or across sessions in the number of trials to criterion.

Confidence ranges and stages of learning. Since we wish to infer subjects' expectancies for "response" CVCs from their confidence ratings, it is necessary to provide evidence that the confidence ratings were valid. If the ratings actually did reflect subjects' knowledge about the paired-associates, relatively high numerical ratings should have been concurrent with relatively accurate three-letter responses. Figure 2b shows that the percentage of correct responses increased with numerically increasing confidence ratings for each subject. Furthermore, the incidence of ratings in the four confidence ranges should have changed as learning progressed. As subjects changed from consistently responding incorrectly to consistently responding correctly to a given "stimulus" CVC, their confidence should have shifted systematically along the scale from

numerically low to numerically high ratings. To investigate this possibility, we divided all presentations of each CVC pair to each subject into three "stages" of learning: (1) trials prior to the first correct response for the pair, (2) trials from the first correct response until the last incorrect response, and (3) trials following the last incorrect response (pairs which were always responded to correctly after the first correct response contributed no trials to stage 2). Table 1 shows, at each stage of learning, the percentage of ratings in each of the four confidence ranges averaged over subjects and CVC pairs. Before responding correctly to

-----  
Insert Table 1 About Here  
-----

a given CVC, subjects tended to indicate that they were wrong; when consistently responding correctly, they tended to indicate that they were right; when responding to a CVC pair with inconsistent accuracy, their ratings were more evenly distributed.

Thus subjects' confidence ratings appear to be a valid index of their knowledge. It is reasonable to assume, therefore, that when subjects indicated that they were "probably right" or "certainly right," they would have expected the "response" CVC to inform them that their three-letter response was correct; conversely, when subjects indicated that they were "probably wrong" or "certainly wrong," they would have expected the "response" CVC to inform them that their three-letter response was incorrect.

#### Average ERPs

Counting task. In Figure 3a the ERPs which were elicited by counted and uncounted CVCs are superimposed. These ERPs have been grand-averaged over subjects and blocks of trials. Two positive-going waves with different

scalp distributions are prominent. One (P280) is larger at the central and

-----  
Insert Figure 3 About Here  
-----

frontal sites and appears equally in the ERPs elicited by the counted and uncounted CVCs. The later positivity (400 to 700 msec after CVC onset) has a centro-parietal maximum and is apparent only in the ERPs elicited by the counted CVCs. This difference in late positivity was observed in each subject's ERPs (Figure 3b). Since the probability of the counted CVC was 16.7% and that of the uncounted CVCs combined was 83.3%, this late positivity seems to be the centro-parietal  $P_{300}^2$  that is elicited by task-relevant, rare events (see review by Donchin, et al., 1978).

As is typically the case, there were individual differences in the scalp distribution of  $P_{300}$ . For comparison with the ERPs from the paired-associate task, these scalp distributions were expressed as percentages of maximum base-to-peak amplitude and are presented in Table 2.

-----  
Insert Table 2 About Here  
-----

Paired-associate task. For each subject, the ERPs elicited by the "stimulus" and the "response" CVCs were each averaged separately for eight categories of trials (ratings in each of the four confidence ranges by two trial outcomes). By necessity, the number of trials in these various categories differed markedly (see Figure 2). No subject had enough trials in the "certainly wrong"-correct category to form a reliable average ERP.

"Stimulus" and "response" CVC ERPs from the seven remaining categories (grand-averaged over subjects at each of the four scalp sites) are



superimposed in Figure 4. A P280 wave, similar to that appearing in the

-----  
Insert Figure 4 About Here  
-----

counting task ERPs, is seen in both the "stimulus" and "response" CVC ERPs. There were no consistent differences in either the latency or amplitude of this wave among the seven categories. The "stimulus" CVC ERPs display relatively little late positivity and, in contrast to the report of Peters, Billinger, & Knott (1977), did not vary systematically in base-to-peak amplitude among the seven categories. The fact that only the "response" CVCs elicited a sizable  $P_{300}$  is consistent with the finding of Rohrbaugh, Donchin and Eriksen (1974) that only the second of a pair of task-relevant stimuli elicited a  $P_{300}$ .

In the "response" CVC waveforms, however, a substantial late positivity with a central-parietal maximum is apparent. Moreover, there was considerable variability in both the amplitude and peak latency of this late positivity among the categories. On correct trials, the positivity was largest when the rating was in the "probably wrong" range and decreased with increasing confidence that the three-letter response was correct. On incorrect trials, it was larger for "certainly right" and "probably right" ratings and decreased with increasing confidence that the three-letter response was incorrect. These trends were pronounced to the extent that at the "probably wrong" confidence level a larger amplitude late positivity was elicited by the "response" CVC on correct trials than on incorrect trials; whereas, at both "probably right" and "certainly right" levels of confidence a larger late positivity was elicited on incorrect trials than on correct trials. That these trends were consistent across subjects is shown in Figure 5, in which the ERPs elicited by correct and incorrect "response"

CVCs are superimposed, for each subject, at the different confidence levels.

Particularly striking in Figures 4 and 5 is the breadth and sometimes

-----  
Insert Figure 5 About Here  
-----

multi-peaked form of the late positivity. It is possible that these average ERPs reflect a sharper-peaked  $P_{300}$  (such as that seen in the counting task) that varied considerably in latency from trial to trial. But it is also possible that the late positivity in the paired-associate ERPs is composed of multiple positive ERP components (see Friedman, Vaughan, & Erlenneyer-Kimling, 1978; Goodin, Squires, Henderson, & Starr, 1978; Roth, Ford, & Kopell, 1978; Stuss & Picton, 1978). Inspection of individual subject's average waveforms across scalp sites failed to reveal any consistent differences in the scalp distribution of either the various peaks in the late positivity, or of the peak positivity among the seven categories. Moreover, individual differences in the scalp distribution of the late positivity in the "response" CVC ERPs (Table II) conformed remarkably to those seen in the counting task (product moment correlation of the base-to-peak amplitudes at corresponding scalp sites was 0.87). Thus we found no indication that the broad late positivities in the paired-associate ERPs reflect anything but a  $P_{300}$  that varied in latency from trial to trial.

Mean base-to-peak amplitudes of these  $P_{300}$ s in the average ERPs from Cz are presented in Table III. Repeated-measures analysis of variance of

-----  
Insert Table 3 About Here  
-----

these base-to-peak amplitudes (6 subjects with repeated measures on 2 trial outcomes X 3 confidence ranges--the "certainly wrong"-incorrect category was

excluded) indicated that the Confidence range X Trial outcome interaction was statistically significant [ $F(2,10) = 19.4$ ;  $p < .001$ ]. When only the "probably wrong" and "probably right" data were analyzed, the Confidence X Outcome interaction remained significant [ $F(1,5) = 13.9$ ;  $p < .05$ ]. A measure of area under the curve (the sum of the digitized voltages between 350-950 msec after CVC onset) yielded similar results.

Latency-adjusted  $\overline{P300}$  amplitude. As stated above, it is possible that the broad  $\overline{P300}$ s in the average ERPs may not have been representative of the waveshape on single trials. It is necessary, therefore, to assess the extent to which the apparent amplitude differences observed in the average ERPs might be due to differences in the latency variability of  $\overline{P300}$  among the single trials which constituted the various averages. To address this question we latency-adjusted our waveforms using the adaptive method described by Woody (1967).<sup>3</sup> Analyses were done on the single-trial ERPs recorded from Cz, after they were pre-processed with the low-pass digital filter mentioned before. To examine the ERP epoch which contained  $\overline{P300}$ , the digitized voltages 400-950 msec after CVC onset were analyzed. For comparison with these results, analyses were also performed on an epoch (850-1500 msec after CVC onset) that presumably contained only background EEG "noise."

The latency-adjusted average ERPs which resulted from analyses of the  $\overline{P300}$  epoch showed slightly sharper  $\overline{P300}$ s than did the unadjusted averages. Mean amplitudes of these latency-adjusted peaks, measured on each trial relative to the unadjusted pre-stimulus baseline, are shown for each category in Table III. An analysis of variance confirmed that the latency-adjusted ERPs manifested the interaction of Confidence range X Trial outcome [ $F(2,10) = 15.2$ ;  $p < .001$ ]. The distributions of latencies chosen by these analyses had consistently smaller standard deviations [ $F(1,5) =$

117.7;  $p < .001$ ] than did the distributions of latencies chosen by the "noise" epoch analyses. This finding indicates that the Woody analyses of the epoch containing  $P300$  detected a latency-varying ERP component, and not simply randomly occurring peaks in the background EEG (see Harris & Woody, 1969).<sup>4</sup> Thus the ERP amplitude differences we observed cannot be attributed to differences in the single-trial variability of  $P300$  latency. Nor can they be attributed to different mixtures of two kinds of trials (for example, trials with and without a  $P300$  or trials with small versus large  $P300$ s). Distributions of single-trial, latency-adjusted amplitudes were examined for each subject and paired-associate category. These distributions, summed over subjects after adjustment for individual differences in amplitude, are presented in Figure 6. Bimodal distributions

-----  
Insert Figure 6 About Here  
-----

in the categories having large mean  $P300$ s would have suggested a mixture of non-homogeneous waveforms. Instead, the  $P300$  distributions appear to reflect relatively uniform single-trial differences in  $P300$  among the various categories.

Finally, since both confidence ratings (Table 1) and trial outcome varied with stages of learning, could some variable related to these stages (or to time on task) account for the apparent effect of the interaction of confidence and trial outcome on  $P300$ ? Figure 7 shows mean amplitudes of the latency-adjusted averages for combinations of confidence ranges, trial outcomes, and stages of learning. The question here is, when broken down by

-----  
Insert Figure 7 About Here  
-----

stages, do trials of each outcome still show  $P_{300}$  differences related to confidence? Two analyses of variance were performed on the mean latency-adjusted amplitudes--one for correct trials and one for incorrect trials (6 subjects with repeated measures on 3 Confidence ranges X 2 Stages of learning). Both analyses showed statistically significant effects of confidence range (for corrects-- $F(2/10) = 25.6$ ;  $p < .001$ ; for incorrects-- $F(2/10) = 8.0$ ;  $p < .01$ ). The only other effect which reached the  $p < .05$  level of significance was the difference in  $P_{300}$  amplitude between stages for the incorrects [ $F(1/5) = 16.2$ ;  $p < .05$ ]. Thus while there was evidence of an effect due to stages of learning, this variable did not account for the interaction of confidence and outcome on  $P_{300}$ .

#### DISCUSSION

Our data indicate that the amplitude of the  $P_{300}$  elicited by the "response" CVCs is determined by the interaction between a trial's outcome and the subject's expectancy concerning that outcome. Neither confidence by itself, nor whether the "response" CVC confirmed or disconfirmed the subject's three-letter response, accounts for the variance in  $P_{300}$ . Rather,  $P_{300}$  amplitude appears to depend on the degree to which the specific outcome of a given trial was unexpected. The lower the subjective probability assigned to an outcome, the larger the elicited  $P_{300}$ . These data thus strengthen the claim that  $P_{300}$  amplitude is dependent on the subjective probability associated with the ERP-eliciting event.

Our notion of subjective probability implies that subjects apply their

knowledge about a given situation to form differential expectancies (subjective probabilities) for the various events which might occur. These expectancies, being derived from external information that is filtered by subjects' perceptual biases, stored in a fallible memory, and tainted by an individual's predilections, are "subjective" in that they need not accurately reflect the objective probabilities with which events occur. Information processing triggered by the occurrence of an event is affected by the expectancy associated with that event. An aspect of the processing invoked by unexpected events is reflected in  $P_{300}$  amplitude. In the paired-associate task, it seems reasonable to infer the subjective probabilities that were assigned to "response" CVCs from subjects' confidence ratings. The pattern of these ratings suggest that subjects' confidence accurately reflected their knowledge. To the extent that subjects thought they were correct in the choice of their three-letter response, they usually were correct (Figure 2b); and as they learned a list, they indicated more often that they were correct (Table 1).

For the present purposes, it is not necessary to define subjective probabilities rigorously, as one would mathematical probabilities. We need not, for example, require that the subjective probabilities assigned to all events possible in a given situation sum to one. We need only assume that subjects' expectancies form an ordinal scale. Given that one event is more unexpected than a second event, we predict that the  $P_{300}$  elicited by the first will be larger than that elicited by the second. Further, we do not imply that subjects are necessarily aware of the probabilities that are assigned internally to stimuli. In some situations it may be possible for individuals to articulate their expectations or to realize that an occurrence was surprising; however, we associate  $P_{300}$  not with the feeling of surprise, but with the processing of surprising events. <sup>2</sup>

In the paired-associate task the events that were assigned differential expectancies were the trial outcomes--information that was indicated to the subjects by the "response" CVCs. We reasoned that subjects would be surprised by the "response" CVC when they expected to be incorrect but were correct, and when they expected to be correct but were incorrect. Moreover, the extent to which these events were unexpected, and would elicit larger  $P300$ s, would be greater the more confident subjects were in the judgment which was disconfirmed. The results agreed with our predictions. When the "response" CVC informed subjects that their three-letter response was correct, the largest  $P300$  was elicited if subjects had indicated "probably wrong." Successively smaller  $P300$ s were elicited if they had indicated "probably right" and "certainly right." But when the "response" CVC informed subjects that their three-letter response was incorrect, the largest  $P300$  was elicited if they had indicated "certainly right" or "probably right," with successively smaller  $P300$ s when "probably wrong" and "certainly wrong" ratings were given. These trends in  $P300$  amplitude were confirmed by single-trial analyses. Although both confidence and trial outcome varied as learning occurred, stages of learning could not account for the effects of the interaction of these two variables on  $P300$  amplitude (see Figure 7). And since CVC pairs were, overall, presented equally often and were not contingent on the subjects' three-letter responses or on the confidence ratings, the results can not be due to differences in the frequency with which particular "response" CVCs occurred. Thus, consistent with the results of K. Squires et al. (1973),  $P300$  was large to the extent that the confidence rating indicated that subjects' expectancy for the obtained trial outcome was low. Recently, this conclusion was also reached by Campbell, Courschesne, Picton, and K. Squires (1979).

Our results strongly support the suggestion that  $P300$  reflects the

subjective probability for a category of stimuli (see Johnson & Donchin, in press). In one case ("probably wrong"-correct) large  $P300$ s occurred when the "response" CVC matched the syllable which the subject had presumably activated in memory, having just typed it as the three-letter response; but in other cases ("probably right"- and "certainly right"-incorrect) large  $P300$ s occurred when the "response" CVC mismatched the three-letter response. Thus  $P300$  amplitude was not dependent on whether or not the subject had anticipated the particular "response" CVC which occurred. Rather, the important variable was whether or not the category to which the "response" CVC belonged (denoting correct or incorrect trial outcome) was surprising.

The notion that individuals assign subjective probabilities to events that may occur in the future seems necessary given the way people deal with uncertainty (see Sheridan & Ferrell, 1974). The less often an uncertain event occurs, the slower subjects respond to it (see review by Smith, 1968), the less likely they are to acknowledge its occurrence (e.g., Swets, Tanner, & Birdsall, 1961), and the less often they predict that it will occur (e.g., Goodnow, 1955). Much effort has been directed at inferring subjective probabilities from behavioral measures (e.g., Edwards, 1962). In some situations, a normative model provides a reasonable approximation to people's performance in estimating stimulus probabilities and predicting uncertain events (see review by Peterson & Beach, 1967). But systematic biases in subjects' performance reveal that subjective probabilities often do not accurately reflect objective probabilities. Predictions and trial-to-trial estimates of probability are consistently conservative relative to a model of optimal behavior. On the other hand, studies of multi-stage inference have shown subjects to be too extreme in their probabilistic inferences (see review by Slovic, Fischhoff, & Lichtenstein, 1977). Furthermore, there is convincing evidence that people sometimes



disregard information about probabilities and use instead various heuristics in forming judgements (Tversky & Kahneman, 1974). And in a random Bernoulli series, where successive events are by definition independent, RT responses vary systematically with the sequence of preceding events (see review by Kornblum, 1973).

The present results are consistent with a growing body of evidence that indicates that ERPs also reflect the differential processing of unexpected stimuli. This evidence suggests that the less probable an event is believed to be--whether because it is being presented relatively infrequently (see reviews by Donchin, et al., 1978 and Ruchkin & Sutton, 1978b), or because it has not occurred recently in a sequence of events (Duncan-Johnson & Donchin, 1977; Duncan-Johnson & Donchin, 1978; Johnson & Donchin, 1978b; Johnson & Donchin, in press; K. Squires et al., 1976; K. Squires et al., 1977), or as shown by the present data, because the event seems unlikely given the subject's current knowledge of a situation--the larger the  $P_{300}$ . Thus when subjective probability varies,  $P_{300}$  amplitude varies.

The extent to which we can make the converse inference, that events that elicit a larger  $P_{300}$  are less subjectively probable, depends on the extent to which other variables known to systematically influence  $P_{300}$  amplitude operate in a given situation. It has been well established that gradations in the task relevance of an event (Johnson & Donchin, 1978a) modulate  $P_{300}$  amplitude. Indeed, most recent accounts of  $P_{300}$  have found it necessary to postulate more than one construct in order to explain the systematic variance in  $P_{300}$  (Donchin, 1979; Donchin et al., 1978; Ruchkin & Sutton, 1978b; K. Squires et al., 1973; Sutton, 1979).

Whether our view of subjective probability is compatible with earlier explanations of  $P_{300}$  in terms of the resolution of uncertainty and delivery of information (Sutton et al., 1965, 1967) depends on what these latter

terms are taken to mean. Would more uncertainty be resolved (or more information be delivered) by the "response" CVC when subjects did not think that they knew the appropriate paired-associate ("certainly wrong" or "probably wrong") than when they did think that they knew it ("certainly right" or "probably right")? If so, then these constructs do not account for the present results. But if more uncertainty would be resolved or more information delivered on trials having an unexpected outcome, then these conceptualizations seem indistinguishable from that of subjective probability. The importance of the present results is not so much that they argue for the superiority of subjective probability over these other constructs but that they constrain what must be meant by any construct with which one attempts to account for the observed effects on  $P_{300}$ .

Finally, we emphasize that to relate  $P_{300}$  amplitude to subjective probability is to assert that  $P_{300}$  reflects a functional process which is executed differently depending on the subjective probability associated with events. The nature of this process, indeed the functional significance of  $P_{300}$ , remains elusive. At present, some sort of context-updating operation (see Donchin et al., 1978) seems a likely candidate for the process manifested by  $P_{300}$ . Knowing the relationship between  $P_{300}$  and constructs such as subjective probability is useful for integrating past ERP results and for predicting those of future studies. But more important, the relationship suggests the use of  $P_{300}$  as a dependent measure in studies of subjective probability, and may guide the design of experiments directed at elucidating the nature of both the cognitive operations and physiological mechanisms which underlie  $P_{300}$ .

## References

- Audley, R. J. Some observations on theories of choice reaction time: Tutorial review. In S. Kornblum (Ed.), Attention and performance IV. New York: Academic Press, 1973, pp. 509-545.
- Campbell, K. B., Courchesne, E., Picton, T. W., & Squires, K. C. Evoked potential correlates of human information processing. Biological Psychology, 1979, 8, 45-68.
- Courschesne, E., Hillyard, S. A., & Courschesne, R. Y. P3 waves to the discrimination of targets in homogeneous and heterogeneous stimulus sequences. Psychophysiology, 1977, 14, 590-597.
- Donchin, E. Event-related brain potentials: A tool in the study of human information processing. In H. Begleiter (Ed.), Evoked brain potentials and behavior. New York: Plenum Press, 1979, pp. 13-88.
- Donchin, E., Callaway, E., Cooper, R., Desmedt, J. E., Goff, W. R., Hillyard, S. A., & Sutton, S. Publication criteria for studies of evoked potentials (EP) in man. In J. E. Desmedt (Ed.), Attention, voluntary contraction and event-related cerebral potentials. Prog. clin. Neurophysiology, Vol. 1, 1977, Karger, Basel, pp. 1-11.
- Donchin, E., & Cohen, L. Average evoked potentials and intramodality selective attention. Electroencephalography & Clinical Neurophysiology, 1967, 22, 537-546.
- Donchin, E., & Heffley, E. Multivariate analysis of event-related potential data: A tutorial review. In D. Otto (Ed.), Multidisciplinary perspectives in event-related brain potential research. Washington, D.C.: U.S. Government Printing Office, EPA-600/9-77-043, 1978, pp. 555-572.
- Donchin, E., Kubovy, M., Kutas, M., Johnson, R., Jr., & Herning, R. I. Graded changes in evoked response (P300) amplitude as a function of cognitive activity. Perception & Psychophysics, 1973, 14, 319-324.
- Donchin, E., Ritter, W., & McCallum, C. Cognitive psychophysiology: The endogenous components of the ERP. In E. Callaway, P. Tueting, & S. Koslow (Eds.), Brain event-related potentials in man. New York: Academic Press, 1978, pp. 349-412.
- Duncan-Johnson, C. C., & Donchin, E. On quantifying surprise: The variation in event-related potentials with subjective probability. Psychophysiology, 1977, 14, 456-467.
- Duncan-Johnson, C. C., & Donchin, E. Series-based vs. trial-based determinants of expectancy and P300 amplitude. Psychophysiology, 1978, 15, 262.
- Edwards, W. Subjective probabilities inferred from decisions. Psychological Review, 1962, 69, 109-135.
- Falmagne, J. C. Stochastic models for choice reaction time with applications to experimental results. Journal of Mathematical Psychology,

1965, 2, 77-174.

Friedman, D., Hakerem, G., Sutton, S., & Fleiss, J. L. Effect of stimulus uncertainty on pupillary dilation response and the vertex evoked potential. Electroencephalography & Clinical Neurophysiology, 1973, 34, 475-484.

Friedman, D., Simson, R., Ritter, W., & Rapin, I. Cortical evoked potentials elicited by real speech words and human sounds. Electroencephalography & Clinical Neurophysiology, 1975, 38, 13-19.

Friedman, D., Vaughan, H. G., Jr., & Erlenmeyer-Kimling, L. Stimulus and response related components of the late positive complex in visual discrimination tasks. Electroencephalography and Clinical Neurophysiology, 1978, 45, 319-330.

Goodin, D. S., Squires, K. C., Henderson, B. H., & Starr, A. An early event-related cortical Potential. Psychophysiology, 1978, 15, 360-365.

Goodnow, J. J. Determinants of choice-distribution in two-choice situations. American Journal of Psychology, 1955, 68, 106-116.

Grubin, M. L., Bauer, J. A., Jr., & Walker, E. C. T. Alice User's Guide, Alice Associates, 29 Wellesley Ave., Natwick, Mass., 1976.

Harris, E. K., & Woody, C. D. Use of an adaptive filter to characterize signal-noise relationships. Computers in Biomedical Research, 1969, 2, 242-273.

Johnson, R. Jr. & Donchin, E. On how P300 amplitude varies with the utility of the eliciting stimuli. Electroencephalography and Clinical Neurophysiology, 1978, 44, 424-437. (a)

Johnson, R., Jr., & Donchin, E. Subjective probability and P300 amplitude in an unstable world. Psychophysiology, 1978, 15, 262. (b)

Johnson, R., Jr., & Donchin, E. P300 and stimulus categorization: Two plus one is not so different from one plus one. Psychophysiology, in press.

Johnson, R. L., Bitzer, D. L., & Slottaw, H. G. The device characteristics of the plasma display element. IEEE Transactions on Electron Devices, 1971, 18, 642-649.

Kornblum, S. Sequential effects in choice reaction time: A tutorial review. In S. Kornblum (Ed.), Attention and performance IV, New York: Academic Press, 1973, pp. 259-288.

Kutas, M., & Donchin, E. Variations in the latency of P300 as a function of variations in semantic categorizations. In D. Otto (Ed.), Multidisciplinary perspectives in event-related brain potential research. Washington, D.C.: U.S. Government Printing Office, EPA-600/9-77-043, 1978, pp. 198-201.

Kutas, M., McCarthy, G., & Donchin, E. Augmenting mental chronometry: The P300 as a measure of stimulus evaluation time. Science, 1977, 197, 792-795.

Laming, D. P. J. Subjective probability in choice-reaction experiments. Journal of Mathematical Psychology, 1969, 6, 81-120.

Nie, N. H., Hull, C. H., Jenkins, J. G., Steinbrenner, K., & Bent, D. H. SPSS: Statistical Package for the Social Sciences, New York: McGraw-Hill, 1975.

Noble, C. Measurements of association value (a), rated associations (a'), and scaled meaningfulness (m') for the 2100 CVC combinations of the English alphabet. Psychological Reports, 1961, 8, 487-521.

Peters, J. F., Billinger, T. W., & Knott, J. R. Event related potentials of brain (CNV and P300) in a paired associate learning paradigm. Psychophysiology, 1977, 14, 579-585.

Peterson, C. R., & Beach, L. R. Man as an intuitive statistician. Psychological Bulletin, 1967, 68, 29-46.

Ritter, W., Simson, R., & Vaughan, H. G., Jr. Association cortex potentials and reaction time in auditory discrimination. Electroencephalography & Clinical Neurophysiology, 1972, 33, 547-555.

Rohrbaugh, J. W., Donchin, E., & Eriksen, C. W. Decision making and the P300 component of the cortical evoked response. Perception & Psychophysics, 1974, 15, 368-374.

Roth, W. T., Ford, J. M., & Kopell, B. S. Long-latency evoked potentials and reaction time. Psychophysiology, 1978, 15, 17-23.

Ruchkin, D. S., & Glaser, E. M. Some simple digital filters for examination of CNV and P300 waveforms on a single trial basis. In D. Otto (Ed.), Multidisciplinary perspectives in event-related brain potential research. Washington, D.C.: U.S. Government Printing Office, EPA-600/9-77-043, 1978, pp. 579-581.

Ruchkin, D. S., & Sutton, S. Emitted P300 potentials and temporal uncertainties. Electroencephalography & Clinical Neurophysiology, 1978, 45, 268-277. (a)

Ruchkin, D. S., & Sutton, S. Equivocation and P300 amplitude. In D. Otto (Ed.), Multidisciplinary perspectives in event-related brain potential research. Washington, D.C.: U.S. Government Printing Office, EPA-600/9-77-043, 1978, pp. 175-177. (b)

Sheridan, T. B., & Ferrell, W. R. Man-machine systems: Information, control, and decision models of human performance. Cambridge, Mass.: MIT Press, 1974.

Slovic, P., Fischhoff, B., & Lichtenstein, S. Behavioral decision theory. Annual Review of Psychology, 1977, 28, 1-39.

Smith, E. E. Choice reaction time--An analysis of the major theoretical positions. Psychological Bulletin, 1968, 69, 77-110.

Smith, S., & Sherwood, B. A. Educational uses of the Plato computer system. Science, 1976, 192, 344-352.

Squires, K. C., Donchin, E., Herning, R. I., & McCarthy, G. On the

influence of task relevance and stimulus probability on event-related potential components. Electroencephalography & Clinical Neurophysiology, 1977, 42, 1-14.

Squires, K. C., Hillyard, S. A., & Lindsay, P. H. Cortical potentials evoked by confirming and disconfirming feedback following an auditory discrimination. Perception & Psychophysics, 1973, 13, 25-31.

Squires, K., Petuchowski, S., Wickens, C., & Donchin, E. The effects of stimulus sequence on event related potentials: A comparison of visual and auditory sequences. Perception & Psychophysics, 1977, 22, 31-40.

Squires, K. C., Wickens, C., Squires, N. K., & Donchin, E. The effect of stimulus sequence on the waveform of the cortical event-related potential. Science, 1976, 193, 1142-1146.

Squires, N. K., Donchin, E., Squires, K. C., & Grossberg, S. Risensory stimulation: Inferring decision-related processes from the P300 component. Journal of Experimental Psychology: Human Perception & Performance, 1977, 3, 299-315.

Stuss, D. T., & Picton, T. W. Neurophysiological correlates of human concept formation. Behavioral Biology, 1978, 23, 135-162.

Sutton, S. P300--Thirteen years later. In H. Begleiter (Ed.), Evoked brain potentials and behavior. New York: Plenum Press, 1979, pp. 107-126.

Sutton, S., Braren, M., Zubin, J., & John, E. R. Evoked-potential correlates of stimulus uncertainty. Science, 1965, 150, 1187-1188.

Sutton, S., Tueting, P., Zubin, J., & John, E. R. Information delivery and the sensory evoked potential. Science, 1967, 155, 1436-1439.

Swets, J. A., Tanner, W. P., Jr., & Birdsall, T. G. Decision processes in perception. Psychological Review, 1961, 68, 301-340.

Tueting, P., Sutton, S., & Zubin, J. Quantitative evoked potential correlates of the probability of events. Psychophysiology, 1970, 7, 385-394.

Tversky, A., & Kahneman, D. Judgment under uncertainty: Heuristics and biases. Science, 1974, 185, 1124-1131.

Woody, C. D. Characterization of an adaptive filter for the analysis of variable latency neuroelectric signals. Medical & Biological Engineering, 1967, 5, 539-553.

Footnotes

<sup>1</sup> Many of the 101 points on the confidence scale were not used by a given subject, while other points were used quite often. Some points, by themselves, were used more than 4% of the time. Points 0 and 100 were the ratings used most often by all subjects. Note that the structure of the task predisposed a large proportion of ratings to the upper end of the scale. Since CVC pairs were presented at random until all were learned, subjects received many presentations of pairs which they already knew before they received a sufficient number of presentations of the pairs which they didn't yet know.

<sup>2</sup> For referring to peaks in the present ERPs, we have adopted the notation suggested by Donchin, Callaway, Cooper, Desmedt, Goff, Hillyard, and Sutton, 1977. The positive going wave which occurs at a mean latency of 280 msec is denoted P280. The late positivity, thought to be the same entity that in some previous experiments occurred at 300 msec, but which here occurs at a much longer latency, is denoted the P $\overline{300}$ .

<sup>3</sup> The Woody procedure, which has been used previously in ERP work (Kutas, et al., 1977; Ruchkin & Sutton, 1978a), involves calculating the cross-correlation function between each single-trial waveform and a template of the ERP signal which varies in latency. The lag at which the maximal cross-correlation (or, as used here, cross-covariance) occurs is assumed to be the latency of the signal on that trial. The single-trial ERPs can then be shifted relative to each other to time-lock on the signal, and a latency-adjusted average can be computed. We used two different approaches to derive templates of the latency-varying signal. First, templates were derived by an iterative procedure (Woody, 1967) whereby the latency-adjusted average of one iteration served as the template for the next iteration, proceeding until the template stabilized. These analyses, since they were

done for each paired-associate category and subject separately, were sensitive to any differences which might have existed in the waveshape of the latency-varying ERP component among the various categories. But since there were fewer trials in some categories than others, the reliability of the various derived templates might have differed systematically across categories. Therefore, to derive a single template which was applicable to each of the seven paired-associate categories, we took advantage of evidence (presented in the text) that the late positivities seen in both the counting and paired-associate tasks were composed of the same component-- $P_{300}$ . Thus, as a second approach, we latency-adjusted each subject's counted CVC ERPs and used this average as the template for a one-pass cross-covariance analysis of single ERPs from each of that subject's paired-associate categories. Since both approaches yielded the same pattern of statistically significant results, we report only the results of the iterative analyses. The analyses of the "noise" epochs also followed the iterative approach, but the latency-adjusted ERP elicited by the counted CVC was for each subject used as the template for the first iteration.

<sup>4</sup> This result does not imply that the detection of  $P_{300}$  was equally reliable in all categories. In the categories where  $P_{300}$  amplitude was relatively small, the Woody procedure may have chosen a spurious EEG peak on a larger proportion of trials than in the categories where  $P_{300}$  was relatively large. Consistent with this possibility, the standard deviations of latencies tended to be largest in the categories where  $P_{300}$  amplitude was smallest. For using the Woody procedure to confirm differences among categories in  $P_{300}$  amplitude after adjusting for latency variability, such trends are not problematic. But these trends do make it difficult to draw conclusions about systematic differences in the latency of  $P_{300}$ , since mean latencies could be biased by the proportion of spurious trials chosen in the



various categories. Thus while the waveforms in Figure 5 show apparent differences in  $\overline{P300}$  latency as well as amplitude, whether these latency differences reflect an overall effect of trial outcome or an interaction between outcome and confidence remains obscure.

TABLE I

At each Stage of Learning:  
the Percentage of Trials in each Confidence Range  
(Averaged over subjects)

| Stage of learning                                       | Confidence Range   |                   |                   |                    |
|---|--------------------|-------------------|-------------------|--------------------|
|   | Certainly<br>Wrong | Probably<br>Wrong | Probably<br>Right | Certainly<br>Right |
| Before first<br>time correct                            | 62.5               | 28.9              | 7.3               | 1.3                |
| From first<br>time correct<br>to last time<br>incorrect | 15.8               | 28.5              | 25.4              | 30.4               |
| After last<br>time incorrect                            | .2                 | 5.3               | 14.8              | 79.8               |

TABLE II

Scalp Distribution of  $\overline{P300}$   
in the Two Tasks for each Subject  
(Percent of Maximum Base-to-Peak Amplitude)

| Subject | Counting task           |     |     |    | Paired-associate task   |     |     |    |
|---------|-------------------------|-----|-----|----|-------------------------|-----|-----|----|
|         | Scalp site <sup>a</sup> |     |     |    | Scalp site <sup>b</sup> |     |     |    |
|         | Fz                      | Cz  | Pz  | Oz | Fz                      | Cz  | Pz  | Oz |
| 1       | 46                      | 89  | 100 | 65 | 52                      | 96  | 100 | 59 |
| 2       | 63                      | 100 | 90  | 48 | 71                      | 100 | 84  | 37 |
| 3       | 65                      | 99  | 100 | 63 | 62                      | 100 | 98  | 52 |
| 4       | 48                      | 94  | 100 | 56 | 52                      | 95  | 100 | 56 |
| 5       | 73                      | 100 | 83  | 24 | 78                      | 100 | 78  | 26 |
| 6       | 64                      | 100 | 90  | 23 | 64                      | 100 | 91  | 24 |

<sup>a</sup>Each measure is based on  $\overline{P300}$  amplitude in the average ERPs elicited by the counted (rare) CVC.

<sup>b</sup>Each measure is based on the mean  $\overline{P300}$  amplitude computed over the seven average ERPs elicited by the "response" CVCs in the various confidence range by trial outcome categories.

TABLE III

$\overline{p300}$  Amplitude Before and After Latency-Adjustment  
for Each Confidence Range by Trial Outcome Category  
( $\mu$ v Base-to-Peak)

| Average ERPs | Confidence Range   |                   |                   |                    |
|--------------|--------------------|-------------------|-------------------|--------------------|
|              | Certainly<br>wrong | Probably<br>wrong | Probably<br>right | Certainly<br>right |
| Unadjusted   |                    |                   |                   |                    |
| Corrects     | --                 | 32                | 26                | 15                 |
| Incorrects   | 19                 | 26                | 33                | 33                 |
| Adjusted     |                    |                   |                   |                    |
| Corrects     | --                 | 36                | 31                | 21                 |
| Incorrects   | 21                 | 28                | 38                | 40                 |

NOTE. These amplitudes are grand-means over subjects.  
There were not enough correct trials in the "certainly wrong"  
confidence range to calculate a valid measure.

## Figure Captions

Figure 1. The events within each trial of the paired-associate learning task.

Figure 2a. The percentage of trials on which each subject used each region of the confidence scale. These regions are groups of ratings that together contain at least 4% of the total trials for a given subject. The bar extending to the right of some graphs indicates that rating 100 was itself a "region" for that subject.

2b. For each region and subject the percentage of trials on which the correct three-letter response was made.

2c. For each subject, the partitioning of the confidence scale that resulted from collapsing regions into the four ranges of confidence that best approximated 0, 33, 67 and 100% correct. The bar extending to the right of some graphs indicates that rating 100 was itself a "range" for that subject.

Figure 3a. Grand-averaged (over subjects) ERPS from the counting task. At each scalp site the ERPs elicited by counted and uncounted CVCs are superimposed.

3b. Digitally filtered average ERPs from Cz for each subject. ERPs elicited by the counted and uncounted CVCs are superimposed.

Figure 4. Grand-averaged (over subjects) ERPs elicited by the "stimulus" and "response" CVCs. Separate averages are shown for trials on which subjects rated their confidence in each of the four ranges and when their three-letter responses were correct and incorrect. There was an insufficient number of trials in the "certainly wrong"-correct category to consider.

Figure 5. For each subject, the digitally-filtered averaged ERPs from

Cz which were elicited by the "response" CVC. ERPs from correct and incorrect trials are superimposed for ratings in each of the four confidence ranges. There was an insufficient number of trials in the "certainly wrong"-correct category to consider.

Figure 6. Distributions of the latency-adjusted, base-to-peak amplitudes of single trials, summed over subjects, for each confidence range and trial outcome category. To adjust for individual differences in amplitude, the mean of each subject's amplitudes over all categories was subtracted from each single-trial amplitude for that subject before the data were combined over subjects.

Figure 7. Grand-means (over subjects) of latency-adjusted  $P_{300}$  amplitude. Measures are from Cz ERPs which were elicited by the "response" CVCs on trials from the various combinations of confidence ranges, stages of learning, and correct and incorrect three-letter responses. When broken down by stages, there was an insufficient number of trials in the "certainly wrong"-correct and "certainly right"-incorrect categories.

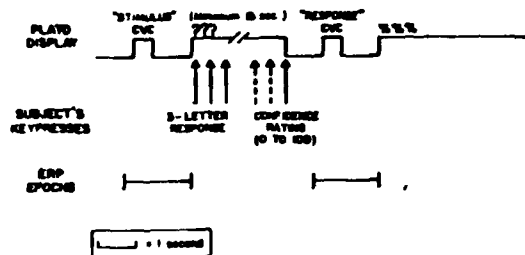


Figure 1

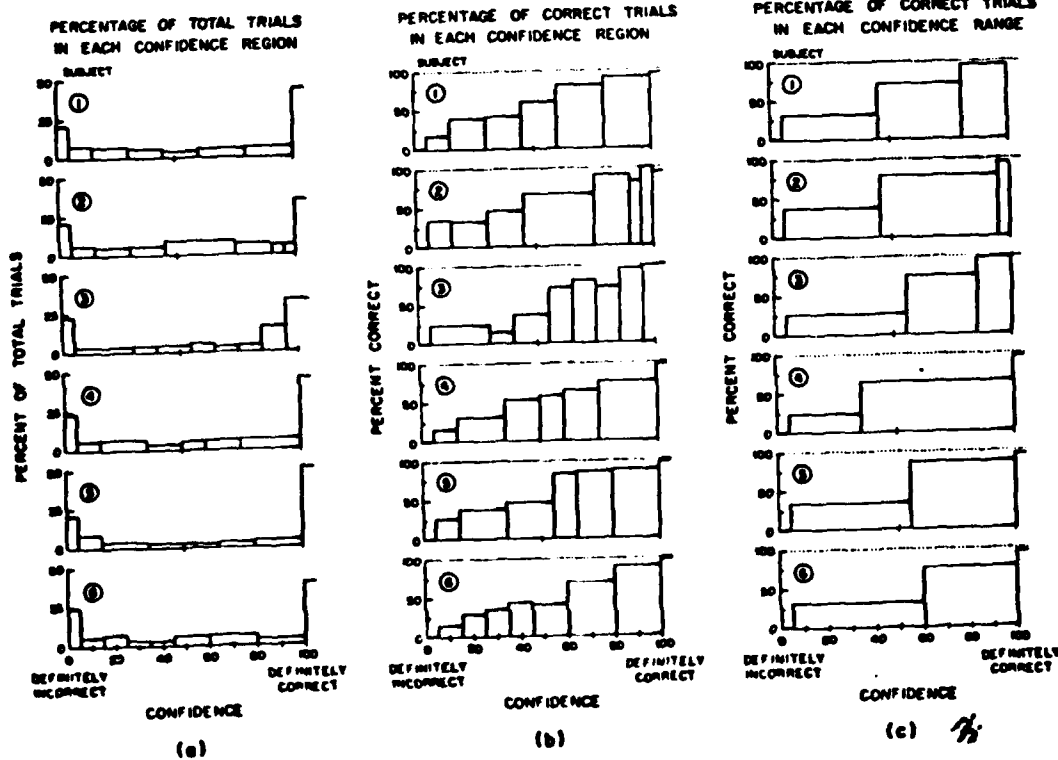


Figure 2

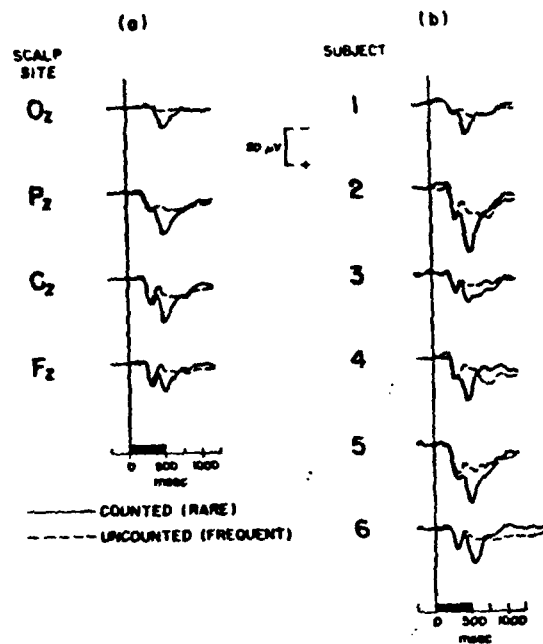


Figure 3

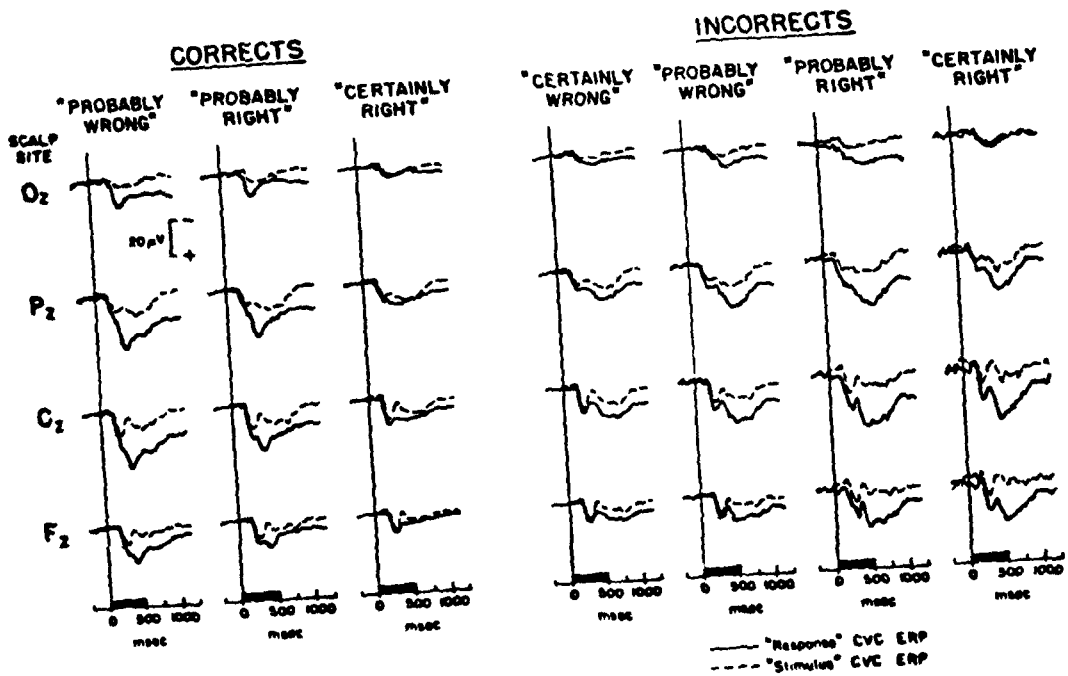
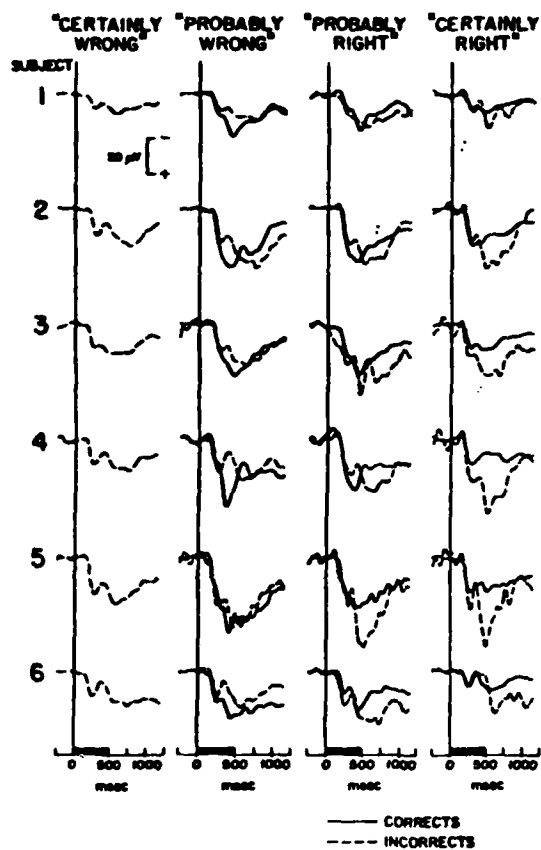


Figure 4





*Figure 5*

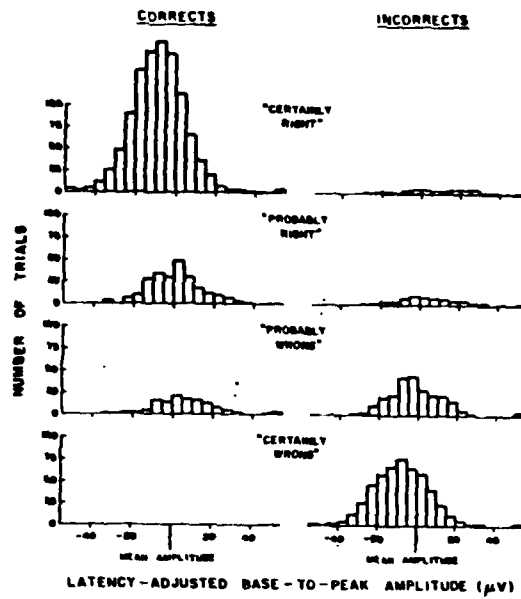


Figure 6

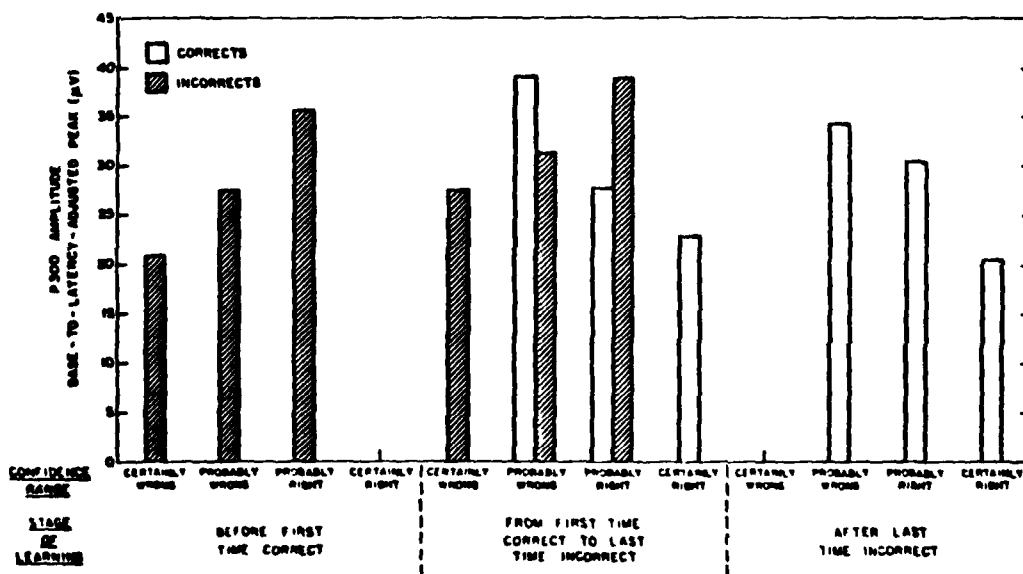


Figure 7

Appendix E

A Metric for Thought:

A Comparison of P300 Latency and Reaction Time

Gregory McCarthy and Emanuel Donchin

Science (in press)

#### Abstract

We confirm that the latency of the P300 component of the human event related potential is determined by processes involved in stimulus evaluation and categorization and is relatively independent of response selection and execution. Stimulus discriminability and stimulus-response compatibility were manipulated independently in an "additive-factors" design. Choice reaction time and P300 latency were obtained simultaneously for each trial. While reaction time was affected by both discriminability and S-R compatibility, P300 latency was affected only by stimulus discriminability.

In his autobiography, Charles Darwin described his fall from the parapet of an old fortification: "...the height was only seven or eight feet. Nevertheless the number of thoughts which passed through my mind during this very short, but sudden and wholly unexpected fall, was astonishing, and seem hardly compatible with what physiologists have, I believe, proved about each thought requiring quite an appreciable amount of time..." (italics added, 1). Darwin was presumably referring to the work of his "friend and contemporary," (2) the Dutch physiologist F. C. Donders who, in 1868, described a technique he used to demonstrate that mental acts have measurable durations. Donders' method was based "on the idea that the time between stimulus and response is occupied by a train of successive processes: each component process begins only when the preceding one has ended" (2). Donders devised a subtractive technique in which "new components of mental action" were interposed in a simple response task. The duration of the added mental component could be determined by subtracting the time required to make a simple response from the time required to make the same response with the additional mental act. From this beginning has developed the study of mental chronometry which seeks to enumerate component mental processes and their characteristics, and to develop models which specify the manner in which these components combine.

Traditional chronometric techniques base inferences about component mental processes on experimental decomposition of the composite reaction time (RT). The analytic power of chronometric techniques would be enhanced if the duration of a subset of the component processes could be recorded concurrently with the composite measure RT. Kutas, McCarthy, and Donchin (3) have suggested that the latency of P300, an event-related brain potential (ERP) recorded in

humans, can serve as such a measure.

There is much evidence that P300 is a manifestation of brain activity invoked during the processing of task-relevant, surprising events (4). The latency of P300 is often positively correlated with RT. However, the correlation between P300 latency and RT can be altered or eliminated by introducing or emphasizing particular factors (3,5). This pattern of correlation suggests that P300 latency is affected by only some of the component processes that contribute to RT. Our hypothesis is that processes concerned with the categorization of stimuli affect P300 latency and RT. Processes of response selection and execution have no effect upon P300 latency. We report here a direct test, and confirmation, of this hypothesis.

We manipulated, in a choice reaction time experiment, two variables whose effects upon RT have been shown to be additive. Thus, we could be reasonably certain that each of the variables was affecting a different processing stage (6). The duration of one stage, which we label stimulus evaluation, was altered by varying the ease with which a target stimulus could be identified (i.e., stimulus 'discriminability'). Response selection was varied by changing the compatibility between the target stimulus and the response required of the subject. As stimulus evaluation is necessary for the categorization of the target, P300 latency should reflect the changes in stimulus discriminability. Changes in stimulus-response compatibility should not affect P300 latency, as the response is selected subsequent to the identification of the target (7). The subject was required to identify which of two target words (RIGHT or LEFT) was embedded in a matrix of characters exposed briefly on a CRT. Four prototypical matrices (8) are illustrated in Figure 1a. In 'noise' (or low discriminability)

trials, the background positions of the matrix were filled with randomly chosen alphabetic characters. In the 'no noise' (or high discriminability) trials, these positions were filled with the "#" symbol.

Subjects indicated the identity of the target word by pressing one of the two response buttons on which the thumb of each hand rested. A cue word, presented in the center of the screen, preceded the exposure of each matrix. The cue SAME indicated that the right button (right thumb) was the appropriate response for the target RIGHT, while the left button was correct for LEFT. The cue OPPOSITE indicated a crossed mapping: the right button (right thumb) was now appropriate for LEFT, and the left button for RIGHT. The stimulus-response mapping, discriminability condition, target word, and position of the target within the matrix were selected randomly on each trial. Each possibility was equally probable, and each was chosen independently of the others.

Stimulus discriminability and S-R compatibility have been demonstrated to have additive effects upon mean RT (9). In a preliminary experiment, we have established that this relationship holds in the specific conditions of our laboratory. The effects of discriminability and S-R compatibility upon mean RT and the percent of correct responses were additive (10).

In the main experiment reaction time and electrophysiological measures were obtained simultaneously (11). Stimulus discriminability and S-R compatibility were again found to have additive effects upon reaction time (12). The mean reaction times for the 'no noise' trials were 624 msec for compatible responses and 716 msec for incompatible responses. For the 'noise' trials, mean RTs of 891 msec and 981 msec

were obtained. Thus the difference between mean RTs due to discriminability was 266 msec, and the difference due to compatibility was 91 msec.

Each artifact-free single trial of EEG data was sorted on the basis of subject, electrode position, target word, discriminability condition, compatibility condition, and correctness of response. The EEG epochs within each sorting bin were averaged. Two sets of averages were obtained, those in which the epochs were aligned by matrix onset, and those in which the epochs were aligned by the subject's response. The response-aligned waveform data will be treated in a later paper. Figure 1b presents ERPs averaged across subjects, and target words, for the midline electrode positions. The matrix elicits an ERP in which a large positive potential is prominent at the parietal electrode site. On the basis of its scalp distribution and latency, we identify this positive potential as the P300 (13).

To quantify the latency of P300, each single trial waveform obtained from the parietal electrode site was low-pass filtered (-3dB at 3.52 Hz) to attenuate EEG activity outside of the bandwidth of P300. The latency of the largest positive peak between 200 and 1500 msec after the onset of the matrix was measured for each trial and used as an estimate of P300 latency. Figure 2 depicts the mean P300 latency estimates and the mean RTs plotted against the experimental variables. The mean P300 latency for the 'no noise,' trials was 589 msec for the compatible response and 617 msec for the incompatible response. For the 'noise' trials, these values were 792 msec and 796 msec. The P300 latency difference of 191 msec due to the discriminability factor was statistically significant ( $F=94.4$ ,  $df=1,12$ ,  $p<.0001$ ). The 16 msec difference associated with the S-R compatibility factor was not



statistically significant ( $F=1.6$ ,  $df=1,12$ ,  $p<.228$ ) The variance of P300 latency was not affected by any experimental variable.

Additional support for our hypothesis is displayed in the rightmost column of Figure 2. The position of the target word within the matrix had a large effect upon mean RT. Targets in either the top or bottom rows were associated with much longer RTs than targets in the middle rows. This effect, however, was restricted to the 'noise' trials (14). According to the additive factors model, this interaction of stimulus discriminability and target position indicates that a common stage is affected by both variables. Therefore, P300 latency should also be affected by target position. This prediction is supported (15) by the similarity of the patterns of RT and P300 latency in Figure 2c (16).

In conclusion, these data confirm the proposition that P300 latency is sensitive to the duration of stimulus evaluation processes, and it is relatively insensitive to response selection processes, while RT is strongly influenced by both. Thus, P300 latency can serve as a metric in the study of mental chronometry. We emphasize that our results do not bear on the nature of the process manifested by P300 (see 4); we only assert that this process is contingent upon stimulus categorization.

Gregory McCarthy

Emanuel Donchin

Cognitive Psychophysiology Laboratory

Department of Psychology

University of Illinois

Champaign, Illinois 61820

Footnotes

- 1). C. Darwin, Autobiography, F. Darwin, Ed. (Schuman, New York, 1950), p. 16.
- 2). F.C. Donders, reprinted in Acta Psychologica, 30, Attention and Performance II, W.G. Koster, Ed. (North-Holland Publishing, Amsterdam, 1969), pp. 412-431.
- 3). M. Kutas, G. McCarthy, E. Donchin, Science, 197, 792 (1977).
- 4). The P300 was first described in S. Sutton, M. Braren, J. Zubin, E.R. John, Science, 150, 1965, 1187-1188. For a contemporary review of P300 see E. Donchin, in Evoked potentials in psychiatry, H. Begleiter, Ed. (Plenum Press, New York, 1979).
- 5). G. McCarthy and E. Donchin, in Brain function in old age, F. Hoffmeister and C. Muller, Eds. (Springer-Verlag, New York, 1979), pp. 318-335; C. Duncan-Johnson, unpublished doctoral dissertation, University of Illinois, 1979. W. Ritter, R. Simson, H.G. Vaughan, Jr., Electroencephalogr. Clin. Neurophysiol., 33, 547 (1972); H.K. Squires, E. Donchin, K.C. Squires, S. Grossberg, J. Exp. Psychol. 3, 299 (1977), J.M. Ford, W.T. Roth, R.C. Mohs, W.F. Hopkins, B.S. Kopell, Electroencephalogr. Clin. Neurophysiol., 47, 1979, 450-459.
- 6). S. Sternberg, in Acta Psychologica, 30, Attention and Performance II, W.G. Koster, Ed. (North-Holland Publishing, 1969), pp. 276-315.
- 7). These assumptions reflect the serial-successive progression of processing stages inherent in the additive-factors model. Other models predict that all processes are active concurrently, although contingent.

upon each other's output. See, for example, J. L. McClelland, Psychological Review, 86, 287, 1979.

8). Each matrix was composed of 4 rows and 6 columns of characters arranged as a square which subtended approximately 2.5 degrees. One target word was presented on each trial, written horizontally, and appearing with equal probability in any of the four rows. The starting column of the target word was also randomly chosen and varied among columns 1 and 2 for RIGHT and columns 1, 2, and 3 for LEFT.

9). I. Bierderman and R. Kaplan, J. Exp. Psychol. 86, 434 (1970); H.W. Frowein and A.F. Sanders, Bulletin of the Psychonomic Society, 12, 106 (1978); S.P. Shwartz, J.R. Pomerantz, H.E. Egeth, J. Exp. Psychol. 3, 402 (1977); Cf. P.M. Rabbitt, Psychonomic Science, 7, 419 (1967).

10). See G. McCarthy, unpublished doctoral dissertation, University of Illinois, 1980.

11). Fifteen male students (right-handed, ages 19-32 years) participated. The matrix was exposed for 400 msec. The cue-to-matrix onset interval was 1000 msec with the cue's exposure duration set for 750 msec of the interval. The scalp EEG was recorded from six Ag/AgCl scalp electrodes (Fz, Cz, Pz, Oz, C3, C4 - according to the 10/20 system) referenced to linked mastoids. Electrodes placed above and to the side of the right eye were used to record the electrooculogram (EOG) in bipolar fashion. The EEG was amplified by a Van Gogh polygraph with a 1/2 amplitude upper cutoff of 35 Hz and with a 10 seconds time constant. The EOG was amplified with an upper cutoff of 15 Hz and with a 1 second time constant. Both the EEG and EOG were

digitized at 5 msec per point for a period of 3.5 seconds beginning 50 msec prior to the cue stimulus and continuing until 2450 msec after the onset of the stimulus matrix. These data were stored on digital tape along with a record of the stimulus conditions and reaction time for that trial. 17.8% of the total trials were not used, either because the subject failed to respond within 2000 msec, or because of eye movement artifact in the EEG.

12). The grand mean RT was 805 msec. The mean RT for 'noise' trials was 266 msec longer than for 'no noise' trials ( $df=1,12$ ,  $F=166.6$ ,  $p<.0001$ ) while the mean RT for incompatible S-R mappings was 91 msec longer than for compatible mappings ( $df=1,12$ ,  $F=84.5$ ,  $p<.0001$ ). Equivalent values were obtained when trials marked for eye movement artifacts were included in the analysis. 'Noise' trials were associated with higher RT variance than 'no noise' trials ( $df=1,12$ ,  $F=32.1$ ,  $p<.0001$ ). There was a nonsignificant trend for more RT variance in the incompatible than compatible trials. Subjects performed correctly on 91.7% of the trials.

13). The large positivity seen in the 'no noise' waveforms is probably a composite of two potentials: one maximum in amplitude over the centro-parietal scalp sites and the other maximum in amplitude over the parieto-occipital scalp sites. The former potential we identify as P300. In the 'noise' trials, these potentials are dissociated in time as the latency of the P300 component increases. On some percentage of the trials, the earlier positive component may have been used to estimate P300 latency. As this component appears relatively fixed in latency, these trials would add a fixed component to the distributions of P300 latency. It is unlikely that this affected our conclusions.

The absence of any significant changes in the variances of the P300 latency distributions suggests that it is unlikely that such misreadings occurred more often in some conditions. For more details see (10).

14). The mean RTs obtained for each matrix row (from the top) were 843 msec, 730 msec, 783 msec, and 883 msec ( $df=3,36$ ,  $F=27.1$ ,  $p<.0001$ ). This row effect strongly interacted with stimulus discriminability ( $df=3,36$ ,  $F=23.5$ ,  $p<.0001$ ) as it was not present in the 'no noise' trials.

15). The mean P300 peak latencies for each matrix row were 721 msec, 665 msec, 684 msec, and 748 msec ( $df=3,36$ ,  $F=7.9$ ,  $p<.0004$ ). As for RT, the row effect interacted with stimulus discriminability ( $df=3,36$ ,  $F=11.0$ ,  $p<.0001$ ) and was not present in the 'no noise' trials.

16). For all experimental factors, the change in mean RT is greater than the change in mean P300 latency. This result is readily apparent in the differing slopes of P300 and RT in figure 2. See (10) for a discussion of these differences and their potential relevance to the assumptions underlying the additive-factors model.

17). This research was supported by the Office of Naval Research under contract number N00014-76-C-0002, with funds provided by the Defense Advanced Research Projects Agency, and by the Air Force Office of Scientific Research, Bolling Air Force Base, Washington, D. C., under contract number F49620-79-C-0233, and the Air Force Systems Command, Wright Patterson Air Force Base, Ohio, under contract number F33615-79-C-0512.

18). G. McCarthy is presently at the Neuropsychology Laboratory, Veterans Administration Medical Center, West Haven, Ct. 06516. We wish to thank E. F. Heffley and C. C. Wood for their helpful comments on previous drafts of this paper.

## Figure Legends

Figure 1(a). Four prototypical matrices used in the experiments. One matrix was presented per trial. The target word RIGHT is present in row 2 of (a), the high discriminability matrix, and in row 1 of (c), the low discriminability matrix. Similar relationships are shown for the target word LEFT in (b) and (c). The starting row, column, target word, and discriminability condition were randomly and independently varied on each trial.

(b) Event-related potentials elicited in the task. The recording epoch is 3050 msec, which comprises a 50 msec pre-stimulus baseline, a 1000 msec epoch between cue onset and matrix onset (vertical line), and 2000 msec of activity following matrix onset. The waveforms presented here represent averages across individual subjects and target words by each discriminability condition ('no noise' or 'noise'), compatibility condition ('compatible' or 'incompatible'), and each midline electrode position (Fz, Cz, Pz, Oz - overlapped at the pre-stimulus baseline).

Figure 2. The mean reaction times (thick lines) and P300 latencies obtained from single-trial measurement (thin lines) for each experimental factor. The main effects of discriminability condition is shown in the left panel. The main effects of Stimulus-Response compatibility is shown in the middle panel. The interaction of discriminability and matrix row is depicted in the rightmost panel.

END

DATE  
FILMED

5 81

DTIC